COO-3077-50

Courant Institute of
Mathematical Sciences

AEC Computing and Applied Mathematics Center

The Method of Artificial Compression: I. Shocks and Contact Discontinuities

Amiram Harten

AEC Research and Development Report

Mathematics and Computers

New York University

COO-3077-50
C.2

UNCLASSIFIED

AEC Computing and Applied Mathematics Center
Courant Institute of Mathematical Sciences
New York University

Mathematics and Computers COO-3077-50

THE METHOD OF ARTIFICIAL COMPRESSION: I. SHOCKS AND CONTACT DISCONTINUITIES

Amiram Harten

June 1974

U. S. Atomic Energy Commission
Contract No. AT(11-1)-3077

UNCLASSIFIED

Table of Contents

	Page
Abstract	v
I. Introduction	1
II. Weak Solutions of a Single Conservation Law and Their Numerical Computation	
1. Weak solutions of a single conservation law	4
2. The viscosity method	11
3. Computation of shocks by finite difference methods	16
4. Computation of contact discontinuities by finite difference schemes	29
III. Artificial Compression Methods	
5. General approach	32
6. Corrective-type artificial compression methods	37
IV. The Δ-Artificial Compression Method	
7. The C_Δ algorithm	40
8. Convergence of successive applications of C_Δ	46
9. Existence of steady progressing solutions of L_Δ	55
10. Limitations of the ACM	59
11. Extension of the ACM to multi-dimensional problems	63
V. Hyperbolic system of Conservation Laws -- Weak Solutions and Their Numerical Computation	
12. Weak solutions of a hyperbolic system of conservation laws	70
13. ACM for hyperbolic system of conservation laws	77
14. Numerical results	86
Acknowledgement	90
References	91
Appendix A. Viscous approximation to numerical solution of shocks	94
Appendix B. Higher order linear approximations to a contact discontinuity	97
Appendix C. Antidiffusion methods.	100
Appendix D. A FORTRAN subroutine.	103
Tables	106
Figures.	112

ABSTRACT

This paper presents a new computational method for the calculation of discontinuous weak solutions of hyperbolic systems of conservation laws, and in particular contact discontinuities. Analysis presented in this report shows that the narrowness of the shock transition is due to a global kind of convexity of the flux function. The idea of the "artificial compression method" (ACM) of the present paper is to modify standard finite difference schemes by adding an artificial flux term which has this property. Analysis and numerical tests show that the ACM is capable of producing an arbitrarily narrow transition of shocks and contact discontinuities.

I. Introduction

In this report we discuss the computation by finite difference methods of discontinuous weak solutions of hyperbolic systems of conservation laws, and in particular the computation of contact discontinuities.

Following the artificial viscosity method of Von Neumann and Richtmyer [24], several other methods for shock calculations were developed (see e.g. [18], [3], [9], [10]). Calculations performed with these methods produce approximate solutions in which a shock is spread over a finite number, usually three to five, of meshpoints. This is quite satisfactory for one-dimensional calculations, but for multi-dimensioned problems better resolution of shocks is desired. Recently a new method based on anti-diffusion considerations was developed by Boris and Book [1]. The spread of shocks computed by this method is typically two cells.

A contact discontinuity is a local linear degeneracy of a nonlinear system of conservation laws. The computation of discontinuities of linear equations, as well as of contact discontinuities, exhibits a seemingly paradoxical phenomenon: the spread of the computed discontinuity is proportional to some power of the number of time-steps taken; this is in contrast to the good resolution of shocks which is independent of the number of time-steps taken (see [13], [16] and [17]).

The speed of propagation of a contact discontinuity is equal to some characteristic speed; for the equations of fluid dynamics, this speed is that of the fluid itself. By using Lagrange coordinates it is possible to integrate explicitly the equation describing the propagation along particle paths, and by this avoid the smearing of the contact discontinuity. This works well enough in one dimension; however, attempts for use of this approach in higher dimensions were not successful.

Richtmyer [21] tried to avoid the smearing of contact discontinuities by using "fitting" techniques. It worked well enough for shocks, but became unstable for contact discontinuities.

The antidiffusion method of Boris and Book [1], although unable to stop the smearing of a contact discontinuity, produce results which are superior to those of standard methods. A brief discussion of antidiffusion methods will be given in Appendix C.

A conservation law is of the form

$$u_t + f(u)_x = 0$$

u is called the conserved quantity, f the flux. Analysis presented in this report shows that the narrowness of the shock transition is due to the nonlinear dependence of the flux on the conserved quantities; the relevant property is a global kind of convexity.

The idea of the method, which we call "artificial compression method", is to add an artificial flux term g, which has this property, but is so constructed that the solution of the original conservation law is also a solution of the altered conservation law

$$u_t + [f(u) + g(u)]_x = 0$$

In the present report we describe how to introduce artificial compression to first order difference schemes. This in itself is not practical because first order schemes are very inaccurate. In a subsequent report we will describe how to use the idea of hybridization of Harten and Zwas [10] to extend our method to schemes of higher order of accuracy.

II. Weak Solutions of a Single Conservation Law and Their Numerical Computation

1. Weak Solutions of a Single Conservation Law

A conservation law asserts that the rate of change of the total amount of a substance contained in a fixed interval [a,b] is equal to the flux of that substance across the boundaries of that interval. Denoting the density of that substance by u, and the flux by f, the conservation law is

$$\frac{d}{dt}\int_a^b u\,dx = -\,f\Big|_a^b \tag{1.1}$$

Taking d/dt under the integral sign and observing $f\Big|_a^b = \int_a^b f_x\,dx$ we obtain

$$\int_a^b (u_t + f_x)\,dx = 0 \tag{1.2}$$

At each point where all partial derivatives of u and f are continuous we obtain the differential conservation law

$$u_t + f_x = 0 \tag{1.3a}$$

We consider an initial value problem of finding a solution $u(x,t)$ to (1.3a) such that

$$u(x,0) = \phi(x)\,, \qquad -\infty < x < \infty \tag{1.3b}$$

where $\phi(x)$ is a given function. Let f be some nonlinear function of u. Denoting

(1.4) $$\frac{df}{du} = a(u)$$

we can write (1.3a) in the form

(1.5) $$u_t + a(u)\, u_x = 0$$

which asserts that u is constant along the characteristic curves $x = x(t)$, where

(1.6) $$\frac{dx}{dt} = a(u)$$

The constancy of u along the characteristics combined with (1.6) implies that the characteristics have a constant slope; so they are straight lines. Their slope, however, depends upon the solution and therefore they may intersect, and where they do, no smooth solutions will exist. To get existence in the large, we admit "weak solutions" which satisfy an integral version of (1.3)

(1.7) $$\int_0^\infty \int_{-\infty}^{\infty} \{w_t u + w_x f(u)\}\, dx\, dt + \int_{-\infty}^{\infty} w(x,0)\, \phi(x)\, dx = 0$$

where u and $f(u)$ are integrable functions over every bounded set of the half-plane $t \geq 0$ and (1.7) holds for every smooth test function $w(x,t)$ of compact support.

If u is a piecewise continuous weak solution, then across the line of discontinuity the Rankine-Hugoniot (RH) relation

(1.8) $$f(u_R) - f(u_L) = S(u_R - u_L)$$

holds, where S is the speed of propagation of the discontinuity, and u_L and u_R are the states on the left and on the right of the discontinuity, respectively (see [14]). Let $g(u)$ be the flux function in the coordinate system which moves with the discontinuity

$$g(u) = f(u) - Su \,. \tag{1.9a}$$

The RH relation (1.8) may be written in the form

$$g(u_L) = g(u_R) \equiv C \tag{1.9b}$$

Thus $g(u)$ is continuous across the discontinuity.

The class of all weak solutions is too wide in the sense that there is no uniqueness for the initial value problem. An additional principle is needed for determining a physically relevant solution. Oleinik [19] gave the following characterization of admissible discontinuities (u_L, u_R, S):

$$\frac{f(u) - f(u_L)}{u - u_L} \geq S \geq \frac{f(u) - f(u_R)}{u - u_R} \tag{1.10}$$

for all u between u_L and u_R ; here S is the speed of propagation of the discontinuity, given by the RH condition (1.8):

$$S = \frac{f(u_R) - f(u_L)}{u_R - u_L} \tag{1.11}$$

(1.10) is called the entropy condition, or condition E. As u tends to u_L and u_R , inequality (1.10) becomes

(1.12) $$a(u_L) \geq S \geq a(u_R)$$

The quantities on the left and on the right in inequality (1.10) can be interpreted, according to the RH condition (1.8) as the speed of propagation of discontinuities (u_L, u) and (u, u_R) respectively. We denote these as $s^L(u)$ and $s^R(u)$:

(1.13) $$s^L(u) = \frac{f(u)-f(u_L)}{u-u_L}, \quad s^R(u) = \frac{f(u)-f(u_R)}{u-u_R}$$

In terms of (1.13), the inequality (1.10) can be restated as

(1.14) $$s^L(u) \geq S \geq s^R(u)$$

for all u between u_L and u_R.

Another way of restating inequality (1.10) is in terms of the function $g(u)$ and the constant C defined in (1.9):

(1.15) $$\frac{g(u) - C}{u - u_L} \geq 0 \geq \frac{g(u) - C}{u - u_R}$$

for all u between u_L and u_R.

A discontinuity is called a shock if in inequality (1.15) the sign of strict inequality holds for all u between u_L and u_R. Of course then the strict inequality also holds in the equivalent inequalities (1.10) and (1.14).

A discontinuity is called a contact discontinuity if equality holds identically in (1.15). Using the definition (1.9) of g it follows that $f(u) = Su + \text{const.}$ for all u between u_L and u_R; thus in this range equation (1.3a) is linear.

Suppose u_L and u_R are two states for which

$$a(u_L) < a(u_R) \tag{1.16}$$

The function

$$u(x,t) = \begin{cases} u_L & \text{for } x < St \\ u_R & \text{for } x > St \end{cases}$$

where

$$S = \frac{f(u_R) - f(u_L)}{u_R - u_L}$$

is a weak solution, but one which violates condition (1.12).

In this case the initial value problem (1.3) with the discontinuous initial data

$$u(x,0) = \begin{cases} u_L & \text{for } x < 0 \\ u_R & \text{for } x > 0 \end{cases} \tag{1.17}$$

has a rarefaction wave solution, i.e. a solution of the form

$$u(x,t) = v(x/t) = \begin{cases} u_L & x/t \leq a(u_L) \\ h(x/t) & a(u_L) \leq x/t \leq a(u_R) \\ u_R & x/t \geq a(u_R) \end{cases} \tag{1.18}$$

such that $a(h(\xi)) = \xi$. If $da/du \geq 0$ for all u between u_L and u_R, (1.18) is continuous for all $t > 0$. A rarefaction wave resolves an initial discontinuity. This is the converse of the process of shock formation.

Weak solutions satisfying condition E are uniquely determined by their initial data. B. Keifitz [12] proved the following more general theorem.

Theorem 1.1. Let f be any C^1 function, u and v two weak solutions of (1.3) of which all discontinuities satisfy condition E (1.10). Then $\|u(t)-v(t)\|$ is a decreasing function, when the norm is the L_1 norm with respect to the x variable.

In particular two solutions which are equal at $t = 0$ are equal for all t.

The uniqueness result can be extended to single conservation laws in any number K of space variables

$$u_t + \operatorname{div} f(u) = 0 , \quad f = (f_1(u),\ldots,f_K(u)) \tag{1.19}$$

The discontinuity is a $K-1$ dimensional hypersurface; let ν be the normal to it and denote by u_+ and u_- the states on both sides of the discontinuity; the + sign corresponds to the region into which the normal is pointing. The RH condition is

$$[f(u_+) - f(u_-)]\cdot\nu = S(u_+ - u_-) \tag{1.20}$$

Let $g(u)$ be the following vector function

$$g(u) = f(u) - (Su)\nu \tag{1.21}$$

The RH condition (1.20) can be restated in terms of (1.21) as

$$g(u_+)\cdot\nu = g(u_-)\cdot\nu \equiv C \tag{1.22}$$

Oleinik's condition E is

(1.23) $[f(u)-f(u_-)]\cdot\nu/(u-u_-) \geq S \geq [f(u)-f(u_+)]\cdot\nu/(u-u_+)$

for all u between u_- and u_+. As for the one-dimensional case, the quantities on the left and on the right can be interpreted, according to the RH condition (1.20) as the speed of propagation of discontinuities (u_-,u) and (u,u_+) respectively. We denote these as $s^-(u)$ and $s^+(u)$:

(1.24)
$$s^-(u) = [f(u)-f(u_-)]\cdot\nu/(u-u_-) \ ;$$
$$s^+(u) = [f(u) - f(u_+)]\cdot\nu/(u - u_+) \ .$$

In terms of (1.24), condition E may be restated as

(1.25) $$s^-(u) \geq S \geq s^+(u)$$

for all u between u_- and u_+. Another way of restating (1.23) is in terms of the function $g(u)$ (1.21) and the constant C:

(1.26) $[g(u)\cdot\nu-C]/(u-u_-) \geq 0 \geq [g(u)\cdot\nu-C]/(u-u_+)$

for all u between u_- and u_+.

2. The Viscosity Method.

One approach for determining physically relevant weak solutions is to admit only those which are obtainable as limits of solution $u_\varepsilon(x,t)$, $\varepsilon \downarrow 0$ of the viscous problem

$$\frac{\partial}{\partial t} u_\varepsilon + \frac{\partial}{\partial x} f(u_\varepsilon) = \varepsilon \frac{\partial}{\partial x} [\beta(u_\varepsilon) \frac{\partial}{\partial x} u_\varepsilon] , \tag{2.1}$$

$$\varepsilon > 0, \quad \beta(u) > 0 , \quad t \geq 0$$

(2.1) is a parabolic equation and it follows from the maximum principle that (2.1) has at most one solution. It is also true that a solution u_ε exists for all t, and that as $\varepsilon \to 0$ these solutions converge in the L_1 sense to a limit u. It was proved in [12] that the discontinuities of u satisfy the entropy condition (1.1a); thus the viscosity method provides a way for constructing admissible solutions of (1.3).

Concerning the question whether every shock is obtainable as a limit of solution u_ε of (2.7), we show the following.

Theorem 2.1. Let u_L and u_R be two values, S the speed given by the RH condition (1.8), and assume that the discontinuous function

$$u(x,t) = \begin{cases} u_L & \text{for } x < St \\ u_R & \text{for } x > St \end{cases} \tag{2.2}$$

is a shock, i.e. that in (1.15) the strict inequality holds for all u between u_L and u_R. Then the solution (2.2) is a limit of solutions u_ε of (2.1).

Proof: The solution u given by (2.2) has the form of a wave travelling with a speed S, i.e. is of the form $u(x,t) = v(x-St)$ with

$$v(y) = \begin{cases} u_L & \text{for } y < 0 \\ u_R & \text{for } y > 0 \end{cases}$$

we construct solutions u_ε of (2.1) approximating u of the same form, i.e.

$$u_\varepsilon(x,t) = v_\varepsilon(x-St) \tag{2.3}$$

Substituting this into (2.1) we find that $v_\varepsilon(y)$ has to satisfy the ordinary differential equation

$$-S\, v_{\varepsilon,y} + f(v_\varepsilon)_y = [\beta(v_\varepsilon)\, v_{\varepsilon,y}]_y \tag{2.4}$$

Setting

$$v_\varepsilon(y) = v(\frac{y}{\varepsilon}) \tag{2.5}$$

and substituting it into (2.4) we find that $v(y)$ has to satisfy the ordinary differential equation

$$-S\, v_y + f(v)_y = [\beta(v)\, v_y]_y \tag{2.6}$$

an equation independent of ε. This equation is easily solved; using the function g introduce in (1.9a), we rewrite it as

$$g(v)_y = [\beta(v)\ v_y]_y$$

We integrate this equation; using the constant C introduced in (1.9b) we get

$$g(v) - C = \beta(v)\ v_y \tag{2.7}$$

We have assumed that the discontinuity (u_L, u_R) is a shock; that means strict inequality holds in (1.15), that is, $g(v) - C$ is nonzero for v between u_L and u_R and of the same sign as $u_R - u_L$, also strict inequality holds at the end points, which means that

$$g'(u_L) > 0 > g'(u_R) \tag{2.8}$$

Under these conditions, and because of the positivity of β, we can solve the ordinary differential equation (2.7) by regarding y as a function of v

$$\frac{dy}{dv} = \frac{\beta(v)}{g(v) - C}$$

Let v_0 and v be two values between u_L and u_R; and let $y(v_0) = y_0$ be an arbitrarily prescribed value of y at v_0; integrating between these boundaries we get

$$y(v) = \int_{v_0}^{v} \frac{\beta(w)}{g(w)-C}\, dw + y_0 \tag{2.9a}$$

Clearly, y is a monotone function of v for v between u_L and u_R, and as v approaches u_L or u_R, y tends to $-\infty$ or $+\infty$ respectively, at a logarithmic rate. It follows

that $v(y)$ is defined for all real y, and as y tends to $-\infty$ or $+\infty$, $v(y)$ approaches u_L or u_R respectively, at an exponential rate.

According to (2.3) and (2.5), $v_\varepsilon(x,t) = v(\frac{x-St}{\varepsilon})$ is a solution of (2.1). Such a travelling wave is called a shock profile for the equation (2.1) with viscosity.

Let v_- and v_+ be any two values between u_L and u_R; we denote by $W_\varepsilon(v_-,v_+)$ the length of the y-interval in which $u_\varepsilon(x,t) = v_\varepsilon(y)$ (2.3) assumes values between v_- and v_+. From (2.5) and (2.9a) we get that $v_\varepsilon(y)$ is given implicitly by

$$\frac{y}{\varepsilon} = \int_{v_0}^{v_\varepsilon(y)} \frac{\beta(w)}{g(w)-C}\, dw$$

Let y_+ and y_- be such that $v_\varepsilon(y_-) = v_-$ and $v_\varepsilon(y_+) = v_+$, respectively. Hence $W_\varepsilon(v_-,v_+) = y_+ - y_-$ can be expressed by

$$W_\varepsilon(v_-,v_+) = \varepsilon \int_{v_-}^{v_+} \frac{\beta(w)}{g(w)-C}\, dw \tag{2.9b}$$

Equation (2.9b) expresses an important property of the shock profile: the width of the transition from v_- to v_+, for any v_- and v_+ between u_L and u_R, is of order ε. Oleinik has shown that every weak solution whose discontinuities are shocks is the limit of a sequence of solutions u_ε of (2.1); then solutions u_ε are continuous and the width of the transition region across the shock of u is of order ε.

We have shown in Theorem 2.1 that strict inequality in condition (1.15) is sufficient for the existence of a steady progressing solution for the equation (2.1), i.e. a shock profile. The converse is also true:

Theorem 2.2. If (2.1) has a unique steady progressing shock profile, then strict inequality holds in (1.15).

Proof: Any such solution of (2.1) necessarily satisfies the ordinary differential equation (2.7). v_y in the right-hand side of (2.7) may not change sign because then the left-hand side which is a function of v alone would admit two different values for the same value of v. Therefore v_y, and consequently $g(v)-C$ has a constant sign, namely that of u_R-u_L; thus (1.15) holds. The strict inequality in (1.15) is implied by the uniqueness theorem.

We remark that all the results concerning uniqueness of solutions of (1.3a) under condition E and the convergence of solutions u_ε of (2.1) to these unique solutions, holds without change if $f(u)$ is replaced by the more general function $f(u,x,t)$, though the characteristics are no longer straight lines (see [12]).

3. Computation of Shocks by Finite Difference Methods.

In this section we shall discuss finite difference schemes which are used to compute solutions of specific initial value problems that come up in scientific and technological problems; if it were possible to prove rigorously convergence of finite difference schemes, this would provide a proof of existence and uniqueness of solutions of (1.3).

Let

$$v(x,t+\Delta t) = L\, v(x,t) \tag{3.1}$$

be a finite difference scheme approximating solutions of (1.3). We say that L has the divergence property if its solution satisfies a difference analog of (1.1), namely, for any $J_2 > J_1$

$$\Delta x \cdot \sum_{j=J_1}^{J_2} v_j^{n+1} = \Delta x \cdot \sum_{j=J_1}^{J_2} v_j^{n} - \Delta t \cdot [(\text{flux terms corresponding to } J_2) - (\text{flux terms corresponding to } J_1)] \tag{3.2a}$$

Δt and Δx are the time and space increments and $v_j^n = v(j\,\Delta x,\, n\,\Delta t)$.

A finite difference scheme is said to be in conservation form if it can be written in the following way.

$$v(x,t+\Delta t) = v(x,t) - \lambda[h(x+\tfrac{1}{2}\Delta x) - h(x-\tfrac{1}{2}\Delta x)] \tag{3.2b}$$

where $\lambda = \Delta t/\Delta x$ and $h(x - \frac{1}{2}\Delta x) = h(v_{-k+1}, v_{-k+2}, \ldots, v_k)$, $h(x + \frac{1}{2}\Delta x) = h(v_{-k}, v_{-k-1}, \ldots, v_{k+1})$. In order for (3.2b) to be consistent with (1.3), h must be related to f as follows.

$$h(u,u,\ldots,u) = f(u) \tag{3.3}$$

A finite difference scheme in conservation form, obviously has the divergence property.

Lax and Wendroff [18] proved the following theorem.

Theorem 3.1. Let $v(x,t)$ be a solution of a finite difference scheme in conservation form. If $v(x,t)$ converges boundedly almost everywhere to some function $u(x,t)$ as Δx and Δt tend to zero, then $u(x,t)$ is a weak solution of (1.3).

A necessary condition for convergence of a finite difference scheme is that the rate of propagation of signals in the difference scheme should be at least as large as the true maximum signal speed, i.e.

$$|a|_{max} \leq \frac{\Delta x}{\Delta t} \tag{3.4}$$

where $|a|_{max}$ is the largest value of the characteristic speed at the relevant range of values. (3.4) is called the Courant-Friedrichs-Lewy condition, or the CFL condition.

Among the first order accurate schemes we consider the well known scheme due to Lax [13] (also known as the

Lax-Friedrichs scheme)

$$(3.5)\quad v_j^{n+1} = (L_1(f)\, v^n)_j = \frac{1}{2}\,(v_{j+1}^n + v_{j-1}^n) - \frac{\lambda}{2}[f(v_{j+1}^n) - f(v_{j-1}^n)]$$

where $v_j^n = v(j\,\Delta x,\ n\,\Delta t)$. This scheme is stable under the CFL condition (3.4); it is in conservation form with $h(x+\frac{1}{2}\Delta x) = \frac{1}{2}\,(v_{j+1} - v_j) + \frac{1}{2}\,[f(v_{j+1}) + f(v_j)]$. The Lax-Friedrichs scheme (LF) is monotone (or monotonicity pr serving scheme) in the sense that if v_j^n is a monotonic function of j, so is v_j^{n+1}.

Another first order accurate scheme, which we shall use in the sequel is the "one-sided scheme". In the linear case

$$u_t + a(x)\, u_x = 0$$

it can be written as

$$(3.6)\quad v_j^{n+1} = v_j^n - \lambda \begin{cases} a(x_j)(u_{j+1} - u_j) & \text{if } a(x_j) < 0 \\ a(x_j)(u_j - u_{j-1}) & \text{if } a(x_j) > 0 \end{cases}$$

If $a(x)$ has a constant sign and $\lambda \max_x |a(x)| \le 1$, then (3.6) is L_2-stable provided that $a(x)$ is Lipschitz continuous; it has positive coefficients and therefore it is monotone (see [15]). If $a(x)$ is changing sign, then (3.6) is still L_2-stable but not necessarily monotone. This lack of monotonicity can be remedied by the following modification.

$$(3.7a)\quad v_j^{n+1} = v_j^n - \lambda a^-_{j+1/2}(v_{j+1}^n - v_j^n) - \lambda a^+_{j-1/2}(v_j^n - v_{j-1}^n)\ ,$$

where the notations $a_{j+1/2} = a(x_{j+1/2})$ and

$$a^+ = \max(0,a), \qquad a^- = \min(0,a) \tag{3.7b}$$

have been used. We claim that (3.7a) with λ such that

$$\lambda \max_k |a(x_{k+1/2})| \leq 1 \tag{3.7c}$$

is a monotonicity preserving scheme.

Proof: For the sake of determinacy let us assume that v_j^n is monotone increasing in j, i.e. $v_{j-1}^n \leq v_j^n \leq v_{j+1}^n \leq v_{j+2}^n$. From (3.7a) we get

$$v_{j+1}^{n+1} = a_{j+1/2}^+ v_j^n + (1 + a_{j+3/2}^- - a_{j+1/2}^+) v_{j+1}^n - a_{j+3/2}^- v_{j+2}^n$$

by replacing v_{j+2}^n by v_{j+1}^n

$$v_{j+1}^{n+1} \geq \lambda a_{j+1/2}^+ v_j^n + (1 - \lambda a_{j+1/2}^+) v_{j+1}^n \tag{3.8a}$$

Similarly by replacing v_{j-1}^n by v_j^n in the following expression

$$v_j^{n+1} = a_{j-1/2}^+ v_{j-1}^n + (1 + a_{j+1/2}^- - a_{j-1/2}^+) v_j^n - a_{j+1/2}^- v_{j+1}^n$$

we get

$$v_j^{n+1} \leq (1 + \lambda a_{j+1/2}^-) v_j^n - \lambda a_{j+1/2}^- v_{j+1}^n \tag{3.8b}$$

By subtracting (3.8b) from (3.8a) we get

$$v_{j+1}^{n+1} - v_j^{n+1} \geq [1 - \lambda(a_{j+1/2}^+ - a_{j+1/2}^-)](v_{j+1}^n - v_j^n) . \tag{3.9}$$

But from (3.7c) $\lambda(a^+_{j+1/2}+a^-_{j+1/2}) = \lambda|a_{j+1/2}| \le 1$, and since $v^n_{j+1}-v^n_j \ge 0$, we get from (3.9) that $v^{n+1}_{j+1}-v^{n+1}_j \ge 0$ too. Hence v^{n+1}_j is also a monotone increasing function of j.

Next we discuss an analogous version of the "one-sided scheme" for the nonlinear equation (1.3). Let $\delta(v,w)$ be

$$\delta(v,w) = \begin{cases} 1 \text{ if } & \frac{f(w)-f(v)}{w-v} < 0 \\ 0 \text{ if } & \frac{f(w)-f(v)}{w-v} \ge 0 \end{cases} \tag{3.10}$$

and $\delta^n_{j+1/2} = \delta(v^n_j, v^n_{j+1})$. With these notations the nonlinear one-sided scheme can be written as, with $f^n_j = f(v^n_j)$,

$$v^{n+1}_j = v^n_j - \lambda\,\delta^n_{j+1/2}(f^n_{j+1}-f^n_j) - \lambda(1-\delta^n_{j-1/2})(f^n_j-f^n_{j-1}) \tag{3.11}$$

We claim that (3.11) is in conservation form and is a monotonicity preserving scheme. To see that (3.11) is in conservation form, i.e.

$$v^{n+1}_j = v^n_j - \lambda[h(v^n_{j+1}, v^n_j) - h(v^n_j, v^n_{j-1})] \tag{3.12a}$$

$$h(v,v) = f(v) \tag{13.12b}$$

we compare (3.12a) with (3.11) and get

$$h(v_{j+1}+v_j) - h(v_j, v_{j-1}) = \delta_{j+1/2}(f_{j+1}-f_j) - (1-\delta_{j-1/2}) \cdot (f_j - f_{j-1})$$

We set $v_{j-1} = v_j$; then by (3.12b), $h(v_{j-1}, v_j) = f_j$ and

$$h(v_j, v_{j+1}) = f(v_j) + \delta_{j+1/2}(f_{j+1}-f_j)\,. \tag{3.13}$$

Indeed (3.13) satisfies (3.12b) and

$$h(v_j, v_{j+1}) - h(v_j, v_{j-1}) = f(v_j) + \delta_{j+1/2}(f_{j+1}-f_j) - f(v_{j-1}) - \delta_{j-1/2}(f_j - f_{j-1})\,,$$

as asserted in (3.13).

To see that (3.11) is monotonicity preserving we observe that (3.11) may be written as

$$(3.14)\quad v_j^{n+1} = v_j^n - \lambda\,\delta_{j-1/2}\,a_{j+1/2}(v_{j+1}^n - v_j^n) - \lambda(1-\delta_{j-1/2})\,a_{j-1/2}(v_j^n - v_{j-1}^n)$$

where

$$(3.15)\quad a_{j+1/2} = \begin{cases} \dfrac{f(v_{j+1})-f(v_j)}{v_{j+1}-v_j}, & v_{j+1} \neq v_j \\ 0, & v_{j+1} = v_j \end{cases}$$

with the notation of (3.7b) $\delta_{j-1/2} \cdot a^-_{j-1/2} = a^-_{j-1/2}$, $(1-\delta_{j+1/2})a_{j+1/2} = a^+_{j+1/2}$. Hence (3.14) has the same form as (3.7a); the latter was shown to be monotonicity preserving under the CFL condition

$$(3.16)\qquad \lambda \max_k |a_{k+1/2}| \leq 1$$

Solutions v of the difference scheme (3.5) are approximations to solutions u of the differential equations (1.3); the smaller Δx, the better is the approximation. We are going to analyze how good an approximation v is to u in case u consists of two constant states u_L and u_R connected by a shock travelling at speed S. That is, u is the steady progressing solution (2.2). In the following we shall study steady progressing profiles of the difference scheme, and discuss the width of the transition from u_L to u_R.

We have used the term monotone operator in the sense of monotonicity preserving operator, i.e., if v_j^n is a monotone function of j, so is v_j^{n+1}. Jennings [22] defined a finite difference scheme

$$v_j^{n+1} = H(v_{j+k}^n, v_{j+k-1}^n, \cdots, v_{j-k}^n) \tag{3.17}$$

to be monotone if H is a monotone increasing function of each of its arguments. Clearly, a finite difference operator which is monotone in the sense of (3.17) is also monotonicity preserving.

We say that a finite difference scheme possesses steady progressing profiles travelling at speed S, if for any rational number $\eta = \lambda S$, where $\lambda = \Delta t/\Delta x$ is in the range of the stability condition, there exists an integer P_η such that ηP_η = integer and for all j

$$v_j^{n+P_\eta} = v_{j-\eta P_\eta}^n \tag{3.18a}$$

and for all n

$$\lim_{j\to-\infty} v_j^n = u_L \ , \quad \lim_{j\to+\infty} v_j^n = u_R \ . \tag{3.18b}$$

Next we bring a somewhat more restricted version of Jennings' results [11] about the existence of steady progressing profiles.

Theorem 3.2. Let $f(u)$ (1.3) satisfy the RH relation and condition E across (u_L, u_R), then any finite difference scheme which is monotone and in conservation form, possesses steady progressing profiles. These steady progressing profiles are monotone functions of j and are uniquely determined by prescribing a value $u_0 \in (u_L, u_R)$ at $x = 0$. Furthermore, if $a(u_L) > S > a(u_R)$ and v_j^0 is such that its difference from some steady progressing profile is in L_1, then the sequence v_j^n converges in L_1 to a steady progressing profile.

Next we discuss the converse to Jennings' theorem. For the sake of simplicity we consider stationary solutions, i.e. $S = 0$. (1.15) reads for $S = 0$,

$$\frac{f(u) - f(u_L)}{u - u_L} > 0 > \frac{f(u) - f(u_R)}{u - u_R}$$

For $u_L > u_R$ this is the same as

$$f(u) < f(u_L) = f(u_R) \equiv C \tag{3.19}$$

for all $u \in (u_R, u_L)$.

Theorem 3.3. Let $v^{n+1} = L\, v^n$ be a 3-points monotone finite difference scheme in conservation form, i.e.

$$v_j^{n+1} = v_j^n - \lambda[h(v_{j+1}^n, v_j^n) - h(v_j^n, v_{j-1}^n)] \equiv H(v_{j+1}^n, v_j^n, v_{j-1}^n) \tag{3.20a}$$

$$h(v,v) = f(v)\ , \tag{3.20b}$$

where H is a strictly monotone increasing function of each of its arguments. If L possesses a stationary profile for all $\lambda \neq 0$ in the range of the probability of the stability condition, then (3.19) holds and the stationary solution is monotone.

Proof: Let $V_j(\lambda)$ be a stationary profile, then by (3.20a) it satisfies $V_j = V_j - \lambda[h(V_{j+1},V_j)-h(V_j,V_{j-1})]$. By using (3.20b), we get for all j,

$$(3.21)\quad h(V_{j+1},V_j) = h(V_j,V_{j-1}) = h(V_{j-1},V_{j-2}) = \ldots = f(u_L) = f(u_R) = C .$$

Let U_λ be the set of all the distinct values of $V(\lambda)$ such that $V_j(\lambda) < V_{j-1}(\lambda)$ and/or $V_j(\lambda) > V_{j+1}(\lambda)$. If $V_j < V_{j-1}$ then by the monotonicity of the difference scheme, and by using (3.20b) and (3.21) we get

$$\begin{aligned} V_j = H(V_{j+1},V_j,V_{j-1}) &> H(V_{j+1},V_j,V_j) \\ &= V_j - \lambda[h(V_{j+1},V_j) - h(V_j,V_j)] \\ &= V_j - \lambda C + \lambda f(V_j) . \end{aligned}$$

Hence $f(V_j) < C$. Similarly if $V_j > V_{j+1}$

$$\begin{aligned} V_j = H(V_{j+1},V_j,V_{j-1}) &< H(V_j,V_j,V_{j-1}) \\ &= V_j - \lambda[h(V_j,V_j) - h(V_{j-1},V_j)] \\ &= V_j - \lambda f(V_j) + \lambda C , \end{aligned}$$

and again we conclude $f(V_j) < C$. Thus for all V_j in U_λ

$$f(V_j) < C \tag{3.22}$$

We claim that U_λ consists of all the distinct values of $V(\lambda)$, and that for all V_j in U_λ, $V_{j-1} \geq V_j \geq V_{j+1}$. If this were not the case then there would exist a value V_{j_0} such that $V_{j_0-1} \geq V_{j_0} < V_{j_0+1}$. But using the same kind of estimates as before $V_{j_0-1} \geq V_{j_0}$ implies $f(V_{j_0}) \geq C$, and $V_{j_0} < V_{j_0+1}$ implies $f(V_{j_0}) > C$, which is a contradiction. Hence $V_j(\lambda)$ is a monotone decreasing function of j, and for each value V_j (3.22) holds.

Using the monotonicity of $V_j(\lambda)$ together with the monotonicity of the finite difference scheme, we get $V_j = H(V_{j+1},V_j,V_{j-1}) \geq H(V_{j+1},V_{j+1},V_j) = V_{j+1} - \lambda h(V_{j+1},V_{j+1}) + \lambda h(V_{j+1},V_j) = V_{j+1} - \lambda[f(V_{j+1}) - C]$. Hence

$$|V_{j+1}-V_j| \leq \lambda|f(V_{j+1})-C| \leq \lambda \max_{v\in(U_R,U_L)}|f(v)-C| \tag{3.23}$$

From (3.23) we see that $\lim_{\lambda\to 0}|V_{j+1}(\lambda) - V_j(\lambda)| = 0$, and therefore for any value of $u \in (u_R,u_L)$ there is some $V_j(\lambda)$ which is arbitrarily close. Since (3.22) holds for all λ, we conclude that (3.19) holds for all $u \in (u_R,u_L)$. This completes the proof of Theorem 3.3.

Theorems 3.2 and 3.3 for the monotone difference schemes in conservation form are completely analogous to Theorems 2.1 and 2.2 for the viscous equation

$$w_t + f(w)_x = \varepsilon(\beta(w)w_x)_x$$

This analogy is suggested by observing that for sufficiently smooth solutions (3.20) approximates

$$w_t + f(w)_x = \Delta t[\beta(w,\lambda)w_x]_x \tag{3.24}$$

to a higher degree of accuracy than the degree of accuracy it approximates the original equation (1.3a).

For example, the LF scheme L_1 is a first order accurate approximation to (1.3a) which means that substituting the solution $u(x,t)$ of (1.3a) into (3.5) yields

$$u(x,t+\Delta t) = L_1 u(x,t) + O((\Delta t)^2)$$

Let $W(x,t)$ be a solution of (3.24), and substitute it into (3.5). Carrying out a Taylor expansion of $L_1 w$, we get

$$\begin{aligned} W(x,t+\Delta t) &= \tfrac{1}{2}\,[w(x+\Delta x,t) + w(x-\Delta x,t)] \\ &\quad - \tfrac{\lambda}{2}\,[f(w(x+\Delta x,t)) - f(w(x-\Delta x,t))] \\ &= w(x,t) - \Delta t\, f_x(w(x,t)) + \frac{(\Delta x)^2}{2} w_{xx} + O((\Delta t)^3 \end{aligned} \tag{3.25a}$$

From (3.24) we get

$$\begin{aligned} w(x,t+\Delta t) &= w(x,t) + \Delta t\, w_t + \frac{\Delta t^2}{2} w_{tt} + O((\Delta t)^3) \\ &= w(x,t) + \Delta t\{-f_x + \Delta t[\beta(w,\lambda)w_x]_x\} \\ &\quad + \frac{(\Delta t)^2}{2} f_{tx} + O((\Delta t)^3) \\ &= w(x,t) - \Delta t\, f_x + (\Delta t)^2\{[\beta(w,\lambda) + \tfrac{1}{2}\, a^2(w)\,]w_x\}_x + O((\Delta t)^3). \end{aligned} \tag{3.25b}$$

Comparing (3.25a) with (3.25b) we see that for

$$\beta(w,\lambda) = \frac{1}{2}\,[1/\lambda^2 - a^2(w)] \tag{3.26}$$

$$w(x,t+\Delta t) = L_1 w(x,t) + O((\Delta t)^3)$$

Thus for all sufficiently smooth problems, the LF scheme (3.5) approximates the solution of (3.24) to a higher degree of accuracy than that to which it approximates solutions of (1.3a).

Although the Taylor expansion is not valid for discontinuous solution, the previous statement is nevertheless true. A numerical verification is given in Appendix A; see also linear analysis in Appendix B.

We use this analogy to study the spread W of the computed shock, which we measure by the number of cells occupied by values between u_- and u_+ . By (3.24) and (2.9b) w may be approximated by

$$W(u_-,u_+) \approx \lambda \int_{u_-}^{u_+} \frac{\beta(w,\lambda)}{g(w)-C}\,dw \tag{3.27}$$

In particular, if we use $\beta(w,\lambda)$ given by (3.26) we get the following estimate for the LF scheme

$$W(u_-,u_+) \approx \frac{1}{2\lambda} \int_{u_-}^{u_+} \frac{1-\lambda^2 a^2(w)}{g(w)-C}\,dw \tag{3.28}$$

Note that the spread depends very strongly on the size of $g(w)-C \geq 0$. If $g(w)-C$ is not much larger than zero then the spread might be unacceptably large.

A similar analysis for the one-sided scheme (3.11) shows that $\beta(w,\lambda) = \frac{1}{2}\,|a(w)|\,[1/\lambda - |a(w)|]$ and therefore

$$W(u_-,u_+) \approx \frac{1}{2}\int_{u_-}^{u_+} \frac{|a(w)|\,[1-\lambda(a(w))]}{g(w)-C}\,dw \tag{3.29}$$

For a stationary shock we get from (1.12) that

$$a(u_L) \geq 0 \geq a(u_R) .$$

Thus a(u) can be close to zero for most values of u between u_- and u_+ ; in this particular case the width w (3.29) will be a lot smaller than that of the LF scheme.

We remark without carrying out the analysis that higher order accurate schemes have usually a better resolution of shocks. The second order accurate scheme of Lax and Wendroff [18] (the LW scheme)

$$\begin{aligned} v_j^{n+1} = (L_2 v^n)_j = v_j^n - \frac{\lambda}{2}\,[f(v_{j+1}^n) - f(v_{j-1}^n)] \\ + \frac{\lambda^2}{2}\,\{a(v_{j+1/2}^n)\,[f(v_{j+1}^n) - f(v_j^n)] \\ - a(v_{j-1/2}^n)\,[f(v_j^n) - f(v_{j-1}^n)]\} \end{aligned} \tag{3.30}$$

where $v_{j\pm 1/2} = (1/2)(v_j + v_{j\pm 1})$, has a typical width of 2-4 cells. On the other hand the LW scheme is not monotone; in particular shock profiles are not, but have an overshoot; other names for this phenomenon are post shock oscillations on parasitic waves. Godunov [7] has shown that higher order schemes cannot be monotone and Lax [15] showed that post shock oscillations are unavoidable.

4. Computation of Contact Discontinuities by Finite Difference Schemes

A contact discontinuity is associated with a linear equation

$$u_t + Su_x = 0 \tag{4.1}$$

$S = \text{const.}$ In the linear case there is neither formation of shock waves, nor a resolution of an initial discontinuity by rarefaction waves; discontinuities are present if and only if they were originally present.

Next we consider approximation to solutions of (4.1) by solutions u_ε of the viscous equation

$$u_t + Su_x = \varepsilon u_{xx} \tag{4.2}$$

Equation (4.2), when considered in a system of coordinates which moves with speed S, reduces to the heat equation

$$u_t = \varepsilon\, u_{yy} \tag{4.3a}$$

If discontinuity is present initially

$$u(y,0) = \begin{cases} u_L\,, & y < 0 \\ u_R\,, & y \geq 0 \end{cases} \tag{4.3b}$$

then

$$u_\varepsilon(y,t) = \frac{1}{2}\,(u_R+u_L) + \frac{1}{2}\,(u_R-u_L)\ \mathrm{Erf}\left(\frac{y}{\sqrt{4\varepsilon t}}\right) \tag{4.4}$$

Thus at time t the width of the transition is proportional to $\sqrt{4\varepsilon t}$.

Next we consider solutions of the LF cheme

$$(4.5)\quad v_j^{n+1} = (L_1 v^n)_j = \tfrac{1}{2}(v_{j+1}^n + v_{j-1}^n) - \tfrac{1}{2}\lambda S(v_{j+1}^n - v_{j-1}^n)$$

to the initial value problem (4.1) and (4.3b); these can be expressed explicitly by

$$(4.6)\quad v_j^n = u_L + (u_R - u_L) \sum_{k=\frac{j+n}{2}}^{n} \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

where $\theta = \frac{1}{2}(1-\lambda S)$. The series in the right-hand side of (4.6) is the binomial distribution function. Using De Moivre-Laplace limit theorem we get that (4.6) can be well approximated for n sufficiently large by

$$(4.7)\quad v_j^n \approx \tfrac{1}{2}(u_R+u_L) + \tfrac{1}{2}(u_R-u_L)\ \mathrm{Erf}\left(\frac{x-St}{\Delta x\sqrt{2n(1-\lambda^2 S^2)}}\right)$$

Thus the width of the transition at time $t = n\,\Delta t$ is proportional to

$$(4.8)\qquad W_1 \approx \text{const.}\ \sqrt{n(1-\lambda^2 S^2)}$$

Comparing (4.7) with (4.4) we see that the LF solution can be approximated by the solution of the viscous equation (4.2) with ε given by (3.26).

It will be shown in Appendix B that the width w_R of the transition at $t = n\,\Delta t$ produced by an R-th order accurate finite difference scheme is

$$(4.9)\qquad w_R \approx \text{const.}\ n^{1/(R+1)}\ .$$

We observe that at a given time t, the width is narrower for higher order accurate schemes; yet the width keeps growing with time and might become unacceptably large.

III. Artificial Compression Methods

5. General Approach.

In this section we shall discuss modifications of finite difference schemes which we call artificial compression methods (ACM). When applied to contact discontinuities these schemes produce approximations in which the discontinuity is spread over a few meshes; when applied to shocks these schemes improve their resolution.

Let u be a weak solution of

$$u_t + f(u)_x = 0 \tag{5.1}$$

and let us assume for the sake of simplicity that u contains a single shock. Denote by $u_L(t)$ and $u_R(t)$ the states on the left and on the right, respectively, of this shock. Suppose that at any time t the function $u(x,t)$ does not take on values between $u_L(t)$ and $u_R(t)$.

Let $g = g(u,t)$ be a function with the following properties:

(5.2) $g(u,t) = 0$ for u outside the interval $\big(u_L(t), u_R(t)\big)$.

Then u is also a solution of

$$u_t + f_x + g_x = 0 \tag{5.3}$$

Suppose in addition that g satisfies

$$\frac{g(u,t)}{u_R - u_L} > 0 \tag{5.4}$$

Then g is called an artificial compression function (ACF).

Suppose that the discontinuity of u, regarded as a solution of (5.1) is a shock, i.e. that (1.15) is satisfied. It follows from (5.4) that the discontinuity of u is also a shock when u is regarded as a solution of (5.3). More important, if the discontinuity of u when regarded as the solution of (5.1) is a contact discontinuity, it becomes a shock when u is regarded as a solution of (5.3).

We show now how to apply the method directly when one knows a priori $u_L(t)$ and $u_R(t)$.

Example: Consider the initial value problem

$$u_t + Su_x = 0 \tag{5.5a}$$

$$u(x,0) = \begin{cases} u^L(x) , & x \leq 0 \\ u^R(x) , & x > 0 \end{cases} \tag{5.5b}$$

where $u^L(x)$ and $u^R(x)$ are monotone decreasing functions and $u^L(0_-) = u_L$, $u^R(0_+) = u_R$. As observed in section 4 solving this problem by the LF scheme (4.4) will result in a spread which is proportional to $\sqrt{n}$. We show now how to solve this initial value problem by an artificial compression method; set

$$g(u) = \begin{cases} \alpha(u-u_L)^2(u-u_R)^2/(u_R-u_L)^3 & \text{for } u \in [u_L,u_R] \\ 0 & \text{otherwise} \end{cases} \tag{5.6a}$$

where α is a parameter. Clearly this function has the desired properties (5.2) and (5.4). We solve now

(5.6b) $$v_j^{n+1} = \frac{1}{2}(v_{j+1}^n + v_{j-1}^n) - \frac{1}{2}\lambda\{[Sv_{j+1}^n + g(v_{j+1}^n)] - [Sv_{j-1}^n + g(v_{j-1}^n)]\} .$$

The CFL condition (3.4) for this scheme is

(5.7) $$\lambda \max_{u\in(u_L,u_R)} |S + a(u)| \le 1$$

where $\lambda = \Delta t/\Delta x$ and $a(u) = dg/du$. Accordingly, we take λ to be

(5.8) $$\lambda = \delta/\max_u |S + a(u)| , \qquad 0 < \delta \le 1 .$$

We shall use now the theory developed in section 3 to study the dependence on the parameter α of the width of the discontinuity in solutios of the difference scheme (5.6). Using the estimate (3.28) for the width of the LF scheme, we get the following upper bound for the width $w(\alpha)$ of the computed discontinuity

(5.9) $$W(\alpha) = \frac{1}{\alpha\,\lambda(\alpha)} \int_{u_-}^{u_+} \frac{1-\lambda^2(\alpha)[S+\alpha a_0(u)]^2}{g_0(u)}\, du$$

Substituting (5.8) for $\lambda = \lambda(\alpha)$ in (5.9) and using the fact that within the range of integration u satisfies $0 \le 1 - \lambda^2(\alpha)\,[S + \alpha\, a_0(u)]^2 \le 1$ we get

(5.10) $$W(\alpha) \le \frac{1}{2\delta}\,(S/\alpha + \max_u |a_0(u)|) \int_{u_-}^{u_+} \frac{du}{g_0(u)} .$$

Thus $W(\alpha)$ is a decreasing function of α.

In Table 1, we give the width taken from the numerical

computation of (5.6), with $u_L = 1$, $u_R = 0$, $u_- = 0.9$, $u_+ = 0.1$ and $\delta = 0.9$.

We turn now to the more general case where we do not know u_L and u_R ahead of time, nor the location of the discontinuity. This information needed for the construction of the ACF has to be extracted from the numerical solution itself. We fit the ideas of the ACM into the numerical framework constructing g at the beginning of each time.

We define the numerical ACF as follows: Let $[x_L(t), x_R(t)]$ be the largest interval around the center of a computed discontinuity in which the computed numerical solution $v(x,t)$ varies monotonically in x between $u_L(t)$ and $u_R(t)$; denote $v(x_L(t),t) = v_L(t)$, $v(x_R(t),t) = v_R(t)$. We say that a function $g(v,x)$ is a numerical ACF for this particular discontinuity if

(5.11) $g(v,x) = O((\Delta x)^R)$ outside the interval $[x_L(t), x_R(t)]$

and

(5.12) $\frac{g(v,x)}{u_R - u_L} > 0$ for all v between v_L and v_R.

In order to preserve the accuracy of the finite difference scheme outside the region of the discontinuity, R in (5.11) should be equal to or larger than the order of accuracy of the scheme.

In section 7 we shall prsent a method of constructing an ACF via the numerical gradient, which automatically satisfies (5.11) and (5.12). Yet, particular problems may

provide some explicit information about the discontinuity, which can be used for a more explicit construction of a numerical ACF.

6. Corrective-type Artificial Compression Methods

We have defined an ACM as a modification of a finite difference approximating (1.3a), $v_j^{n+1} = L(f)\, v_j^n$, such that it will approximate an equation

$$u_t + [f(u) + g(u)]_x = 0 \tag{6.1}$$

where $g(u)$ is an ACF. In section 5 we have considered doing so by simply using the same scheme L to solve (6.1) i.e.

$$v_j^{n+1} = L(f+g)v_j^n \ . \tag{6.2}$$

This method, which we shall call a built-in ACM, has an obvious computational disadvantage -- it involves a reduction of the permissible time-step (see (5.8) and Table 1). It seems advisable to separate the artificial compression from the main calculation. This can be done by solving (6.1) by a "split-flux" method: Let L be some standard finite difference scheme in conservation form for solving the original equation (1.3a) in the sense of

$$v_j^{n+1} = (Lv^n)_j = v_j^n - \lambda[h(v_{j+1}^n, v_j^n) - h(v_j^n, v_{j-1}^n)] \tag{6.3a}$$

$$h(v,v) = f(v) \ . \tag{6.3b}$$

Let C be some finite difference scheme in conservation form for solving the equation $u_t + g_x = 0$

$$v_j^{n+1} = (Cv^n)_j = v_j^n - \lambda[\ell(v_{j+1}^n, v_j^n) - \ell(v_j^n, v_{j-1}^n)] \tag{6.4a}$$

$$\ell(v,v) = g(v) \ . \tag{6.4b}$$

We consider the combination of these two schemes $\hat{L} = C\,L$, i.e.

$$w^{n+1} = L\,v^n \tag{6.5a}$$

$$v^{n+1} = C\,w^{n+1} = \hat{L}\,v^n \ . \tag{6.5b}$$

$\hat{L}$ can be written explicitly as

$$v_j^{n+1} = (Lv^n)_j = v_j^n - (k_{j+1/2} - k_{j-1/2}) \tag{6.6a}$$

where

$$\begin{aligned} k_{j+1/2} &= k(v_{j+2}, v_{j+1}, v_j, v_{j-1}) \\ &= h_{j+1/2} + \ell[v_{j+1} - (h_{j+3/2} - h_{j+1/2}), \\ &\qquad v_j - \lambda(h_{j+1/2} - h_{j-1/2})] \end{aligned} \tag{6.6b}$$

$$h_{j+1/2} = h(v_{j+1}, v_j) \ .$$

Hence $\hat{L}$ too is in conservation form. Setting $v_{j+2} = v_{j+1} = v_j = v_{j-1} = v$ in (6.6b) we get

$$k(v,v,v,v) = f(v) + g(v) \tag{6.7}$$

Thus $\hat{L}$ is consistent with (6.1). Furthermore, if both L and C are monotonicity preserving schemes, so is $\hat{L}$.

We observe that $\hat{L}$ is a corrective-type scheme: L smears the discontinuity while propagating it; then the smeared discontinuity is compressed back by applying the artificial compression C. This operator is an approximation to the time evolution under the equation

(6.8) $$u_t + g(u)_x = 0$$

which is a pure compression; for this reason the time-step associated with C is a dummy time-step.

In order to increase the compressibility of L, one may consider $L = C^p L$ with $p > 1$. This scheme, which is again monotone and in conservation form, is consistent with the equation $u_t + [f(u) + p\, g(u)]_x = 0$. It will be seen in the next chapter that $p = 1$ is adequate.

Corrective type ACM has several advantages compared with the built in ACM in addition to the one already mentioned at the beginning of this section: we are at freedom to choose a different scheme for the compression step. As was earlier remarked in section 3, the one-sided scheme (3.11) has a particularly good resolution of a stationary shock (see (3.29)). Therefore (3.11) is a good choice for the artificial compressor.

Another practical advantage of the corrective type ACM is that it can be easily added to an existing computer program in order to sharpen resolution of shocks and prevent the smearing of contact discontinuities; a subroutine for the operation C, and a statement calling this subroutine in the main program is all that is needed.

IV. The Δ-Artificial Compression Method.

7. The C_Δ Algorithm.

In this section we shall show how to construct a numerical ACF, which we call Δ-ACF, without using a priori knowledge about the location and the jump quantities of the discontinuities. Then we shall describe an artificial compression C_Δ based on it. We recall from Section 5 that $g(v,x)$ is a numerical ACF for the discontinuity (u_L,u_R) if for all v in the transition region

$$g(v,x)/(u_R-u_L) > 0 \tag{7.1}$$

and outside the transition region (R = 1 in (5.11))

$$g(v,x) = O(\Delta x) \tag{7.2}$$

We observe that $\Delta_{j+1/2} = v_{j+1} - v_j$ is a good raw material for a numerical ACF: It is $O(\Delta x)$ whenever v is smooth and it has the same sign as u_R-u_L in the transition region. Furthermore, when restricted to a region where v_j is a strictly monotone function of j, $\Delta_{j+1/2}$ (as any other function of j), can be regarded as a function of v, since the correspondence between v and j is 1-1 there.

Let $\{J_m\}_{m=1}^n$ be a partition such that between any two successive points, v_j is a monotone function of j. To each of the monotone regions we assign a function $\tilde{g}$;

for the monotone region $J_L \le j \le J_R$, $\tilde{g}$ is the following

$$\tilde{g}_{j+1/2} = \begin{cases} 0 , & j < J_L \\ v_{j+1} - v_j , & J_L \le j < J_R \\ 0 , & J_R \le j \end{cases} \tag{7.3}$$

It follows immediately from (7.3) that for all j

$$|\tilde{g}_{j+1/2} - \tilde{g}_{j-1/2}| \le |v_{j+1} - 2v_j + v_{j-1}| . \tag{7.4a}$$

Thus if $[J_L, J_R]$ is a region of smoothness of v, then the quantity on the right side of (7.4a) is $O((\Delta x)^2)$, i.e.

$$|\tilde{g}_{j+1/2} - \tilde{g}_{j-1/2}| = O((\Delta x)^2) \tag{7.4b}$$

In such a region, an artificial compressor using this function has the accuracy of a first order scheme. If on the other hand $[J_L, J_R]$ contains a transition region of a discontinuity, then by what was earlier said on $\Delta_{j+1/2} = v_{j+1} - v_j$, it is a numerical ACF, i.e. it satisfies (7.1) and (7.2). In case v_j is not strictly monotone in some isolated points, in order to make $|\tilde{g}|$ nonzero, we increase a minimal value of $|\tilde{g}|$ up to the value of the smaller of its two immediate neighbors. We denote this operation by F and the resulting function by $\hat{g}$. (This operation is illustrated in Figure 1).

$$\hat{g} = F\,\tilde{g} \tag{7.5a}$$

$$\hat{g}_{j+1/2} = \begin{cases} \text{sgn. } \min\{|\tilde{g}_{j-1/2}|, |\tilde{g}_{j+3/2}|\} \\ \qquad \text{if } |\tilde{g}_{j-1/2}| > |\tilde{g}_{j+1/2}| < |\tilde{g}_{j+3/2}| \\ \tilde{g}_{j+1/2} \qquad \text{otherwise} \end{cases} \tag{7.5b}$$

sgn is the sign of $\tilde{g}$, which is constant.

F preserves the sign of $\tilde{g}$, and in addition (see Figure 1),

$$(7.6a) \quad |\hat{g}_{j+1/2} - \hat{g}_{j-1/2}| \le |\tilde{g}_{j+1/2} - \tilde{g}_{j-1/2}| \le |v_{j+1} - 2v_j + v_{j-1}|$$

The last inequality is (7.4a). Thus if $[J_L, J_R]$ is a region of smoothness of v, then

$$(7.6b) \quad |\hat{g}_{j+1/2} - \hat{g}_{j-1/2}| = O((\Delta x)^2) \ .$$

Using $\hat{g}$ instead of $\tilde{g}$ means more compression since according to formula (3.27), the larger $|g|$, the narrower is the transition region.

Next we switch to integer points by

$$(7.7) \quad g_j = \text{sgn} \ \min \{|\hat{g}_{j-1/2}|, |\hat{g}_{j+1/2}|\}.$$

Notice that according to (7.7), $g_j = O(\Delta x)$ at the end of the transition region; this would not be the case if g_j had been defined directly by centered differences as $v_{j+1} - v_{j-1}$, since then the centered difference which is assigned to the end point contains part of the transition itself, which is independent of Δx (see Figure 2).

We make the following observations

$$(7.8) \quad g_{J_L} = g_{J_R} = 0$$

Since $g_{j+1/2}$ has the same sign for all j

$$(7.9a)\quad |g_{j+1} - g_j| = |\min(|\hat{g}_{j+3/2}|, |\hat{g}_{j+1/2}|) - \min(|\hat{g}_{j+1/2}|, |\hat{g}_{j-1/2}|)|$$
$$\leq \max(|\hat{g}_{j+3/2} - \hat{g}_{j+1/2}|, |\hat{g}_{j+1/2} - \hat{g}_{j-1/2}|)$$

Hence if $[J_L, J_R]$ is a region of smoothness of v_j, then by (7.6b),

$$(7.9b)\qquad |g_{j+1} - g_j| = O((\Delta x)^2)$$

To each of the monotone regions $[J_m, J_{m+1}]$, we assign an artificial compressor $C_\Delta^{(m)}$, which is the one-sided difference scheme (3.11) approximating

$$v_t + g_x = 0$$

g is the function defined by (7.3), (7.5) and (7.7) corresponding to this particular monotone region. The operation of the $C_\Delta^{(m)}$ corresponding to the monotone region $[J_m, J_{m+1}]$ is as follows.

$$(7.10a)\quad w_j^{(m)} = (C_\Delta^{(m)} v)_j$$
$$= \begin{cases} v_j - \hat{\lambda}\delta_{j+1/2}(g_{j+1} - g_j) - \hat{\lambda}(1-\delta_{j-1/2})(g_j - g_{j-1}), \\ \qquad\qquad J_m + 1 \leq j \leq J_{m+1} - 1 \\ v_j \qquad\qquad \text{otherwise} \end{cases}$$

where sgn is the sign of g in $[J_m, J_{m+1}]$ and $\delta_{j+1/2}$ is defined by

(7.10b) $$\delta_{j+1/2} = \begin{cases} 1 & \text{if } (g_{j+1}-g_j)\ \text{sgn} < 0 \\ 0 & \text{if } (g_{j+1}-g_j)\ \text{sgn} \geq 0 \end{cases}$$

$C_\Delta^{(m)}$ needs to be applied only for such m that $[J_m, J_{m+1}]$ contains a discontinuity; in most practical problems these monotone regions are easily identified. However, if we cannot locate the relevant monotone regions, then we apply (7.10) to <u>all</u> of them

(7.11) $$w_j = (C\ v)_j = \begin{cases} w_j^{(m)}, & J_m+1 \leq j \leq J_{m+1}-1 \\ v_j, & j = J_m,\ J_{m+1} \end{cases}$$

From (7.10) and (7.9) we get that in the region of smoothness

(7.12) $$w_j = (C\ v)_j = v_j + O((\Delta x)^2)$$

Thus no harm is done to the first order accuracy there.

The analysis of the one-sided scheme (3.11) presented in Section 3 shows that

(i) C_Δ is in conservation form.

(ii) C_Δ is a monotonicity preserving operation provided that

(7.13a) $$\hat{\lambda} \max_j |\gamma_{j+1/2}| \leq 1$$

where

(7.13b) $$\gamma_{j+1/2} = \begin{cases} \dfrac{g_{j+1}-g_j}{v_{j+1}-v_j}, & v_{j+1}-v_j \neq 0 \\ 0, & v_{j+1} = v_j . \end{cases}$$

Thus for $\hat{\lambda}$ satisfying (7.13), w_j is monotone in $[J_m, J_{m+1}]$ for all m; the extremal values of v_j are preserved and we conclude

(7.14) $$\|C_\Delta\|_{max} = 1 .$$

(iii) When applied to smooth v

$$\|C_\Delta v\|_{L_2} \leq (1 + O(\Delta x)) \|v\|_{L_2} \tag{7.15}$$

Next we show that for all j

$$|\gamma_{j+1/2}| \leq 1 \tag{7.16}$$

Proof: Since $\hat{g}_{j+1/2}$ all have the same sign,

$$|g_{j+1}-g_j| = |\min(|\hat{g}_{j+3/2}|, |\hat{g}_{j+1/2}|) - \min(|\hat{g}_{j+1/2}|, |\hat{g}_{j-1/2}|)| \leq |\hat{g}_{j+1/2}|$$

If $|\tilde{g}_{j+1/2}|$ is not a minimum of $|\tilde{g}|$, then

$$|\hat{g}_{j+1/2}| = |\tilde{g}_{j+1/2}| = |v_{j+1}-v_j|$$

and (7.16) follows from (7.13b). In case $|\tilde{g}_{j+1/2}|$ is a minimum of $|\tilde{g}|$, we get from (7.5) that either

$$|\hat{g}_{j+3/2}| \geq |\hat{g}_{j+1/2}| = |\hat{g}_{j-1/2}|$$

or

$$|\hat{g}_{j+3/2}| = |\hat{g}_{j+1/2}| \leq |\hat{g}_{j-1/2}| .$$

In both cases $|g_{j+1}-g_j| = 0$ and therefore $\gamma_{j+1/2} = 0$. This completes the proof that $|\gamma_{j+1/2}| \leq 1$. Thus the CFL condition (7.13) is satisfied by

$$\hat{\lambda} \leq 1 \tag{7.17}$$

A FORTRAN subroutine performing the operation $w^{(m)} = C_\Delta^{(m)} v$ (7.10) is given in Appendix D.

8. Convergence of Successive Applications of C_Δ.

In this section we shall prove that C_Δ defined in Section 7 is indeed a compressor, that is, successive applications of C_Δ to initial data of the form

$$w_j^0 = \begin{cases} v_L , & j \le J_L \\ v_j , & J_L \le j \le J_R \\ v_R , & J_R \le j \end{cases} \tag{8.1a}$$

where v_j is a strictly monotone function of j, and

$$v_{J_L} = v_L , \quad v_{J_R} = v_R \tag{8.1b}$$

transforms it into a solution w^∞ which has a shock-like structure:

$$w_j^\infty = \begin{cases} v_L , & j < J_\infty \\ v_L + \alpha(v_R - v_L) , & j = J_\infty \\ v_R , & j > J_\infty \end{cases} \tag{8.2}$$

$$J_L \le J_\infty \le J_R , \qquad 0 \le \alpha \le 1 .$$

Let w^n be

$$w^n = C_{\Delta,n-1}\, w^{n-1} = \Big(\prod_{k=0}^{n-1} C_{\Delta,k} \Big) w^0 \tag{8.3}$$

where $C_{\Delta,k}$ is defined by

$$w_j^{k+1} = (C_{\Delta,k} w^k)_j = w_j^k - \hat{\lambda}\delta^k_{j+1/2}(g^k_{j+1} - g_j) - \hat{\lambda}(1-\delta^k_{j-1/2})(g^k_j - g^k_{j-1}) . \tag{8.4}$$

g_j^k are defined by (7.3), (7.5) and (7.7) with respect to w^k.

Let $J_{L,n}$ be the largest j such that $w_j^n = v_L$ and let $J_{R,n}$ be the smallest j such that $w_j^n = v_R$. It follows from (7.10a) that $J_{L,n}$ is a nondecreasing function of n, and $J_{R,n}$ is a nonincreasing function of n, and both lie in $[J_L, J_R]$. Thus both sequences are convergent and we denote their limits by

$$\lim_{n\to\infty} J_{L,n} = J_{L,\infty}\ , \qquad \lim_{n\to\infty} J_{R,n} = J_{R,\infty}\ .$$

Lemma 8.1. Let $w^{n+1} = C_{\Delta,n} w^n$ (8.3) and let $\tilde{g}^n_{j+1/2}$, $\hat{g}^n_{j+1/2}$, g^n_j and $\tilde{g}^{n+1}_{j+1/2}$, $\hat{g}^{n+1}_{j+1/2}$, g^{n+1}_j be the quantities defined by (7.3), (7.5) and (7.7) for w^n and w^{n+1} respectively. Then

$$(8.5')\quad |\tilde{g}^{n+1}_{j+1/2}| \ \geq\ \min\left(|\tilde{g}^n_{j-1/2}|, |\tilde{g}^n_{j+1/2}|, |\tilde{g}^n_{j+3/2}|\right)\ .$$

Proof: Multiplying (7.10a) by $\mathrm{sgn} = \mathrm{sign}(v_R - v_L)$, we get

$$(8.6a)\quad \mathrm{sgn}.\ w^{n+1}_{j+1} = \mathrm{sgn}.\ w^n_{j+1} - \hat{\lambda}\delta_{j+3/2}\left(|g^n_{j+2}| - |g^n_{j+1}|\right) - \hat{\lambda}(1-\delta_{j+1/2})\left(|g^n_{j+1}| - |g^n_j|\right)$$

$$(8.6b)\quad \mathrm{sgn}.\ w^{n+1}_j = \mathrm{sgn}.\ w^n_j - \hat{\lambda}\delta_{j+1/2}\left(|g^n_{j+1}| - |g^n_j|\right) - \hat{\lambda}(1-\delta_{j-1/2})\left(|g^n_j| - |g^n_{j-1}|\right)\ .$$

By subtracting (8.6b) from (8.6a) we get

$$(8.7)\quad |\tilde{g}^{n+1}_{j+1/2}| = \mathrm{sgn}\ (w^{n+1}_{j+1} - w^{n+1}_j) = |\tilde{g}^n_{j+1/2}| - \hat{\lambda}(1-2\delta_{j+1/2})\cdot\left(|g^n_{j+1}| - |g^n_j|\right) - \hat{\lambda}\delta_{j+3/2}\left(|g^n_{j+2}| - |g^n_{j+1}|\right) + \hat{\lambda}(1-\delta_{j-1/2})\left(|g^n_j| - |g^n_{j-1}|\right).$$

From the definition (7.10b) of $\delta_{k+1/2}$ it is obvious that

$$\delta_{k+1/2}(|g_{k+1}| - |g_k|) \leq 0$$

Hence the last two terms in the right-hand side of equality (8.7) are positive; therefore

$$|\tilde{g}^{n+1}_{j+1/2}| \geq |\tilde{g}^n_{j+1/2}| - \hat{\lambda}(1-2\delta_{j+1/2})(|g^n_{j+1}|-|g^n_j|)$$

$$= |\tilde{g}^n_{j+1/2}| - \hat{\lambda}\left||g^n_{j+1}|-|g^n_j|\right|$$

Using the definition (7.7) of g_j , we get

$$(8.8)\quad |\tilde{g}^{n+1}_{j+1/2}| \geq |\tilde{g}^n_{j+1/2}| - \hat{\lambda}\Big\{\max[\min(|\hat{g}^n_{j-1/2}|, |\hat{g}^n_{j+1/2}|), \min(|\hat{g}^n_{j+1/2}|, |\hat{g}^n_{j+1/2}|)] - \min(|\hat{g}^n_{j-1/2}|, |\hat{g}^n_{j+1/2}|, |\hat{g}^n_{j+3/2}|)\Big\}$$

$$\geq |\tilde{g}^n_{j+1/2}| - \hat{\lambda}|\hat{g}^n_{j+1/2}| + \hat{\lambda}\min(|\hat{g}^n_{j-1/2}|, |\hat{g}^n_{j+1/2}|, |\hat{g}^n_{j+3/2}|).$$

If $\hat{g}^n_{j+1/2} = \tilde{g}^n_{j+1/2}$, then (8.5) follows from

$$|\tilde{g}^{n+1}_{j+1/2}| \geq (1-\hat{\lambda})|\hat{g}^n_{j+1/2}| + \hat{\lambda}\min(|\hat{g}^n_{j-1/2}|, |\hat{g}^n_{j+1/2}|, |\hat{g}^n_{j+3/2}|)$$

$$\geq \min(|\hat{g}^n_{j-1/2}|, |\hat{g}^n_{j+1/2}|, |\hat{g}^n_{j+3/2}|)$$

$$\geq \min(|\tilde{g}^n_{j-1/2}|, |\tilde{g}^n_{j+1/2}|, |\tilde{g}^n_{j+3/2}|) .$$

On the other hand, $\hat{g}^n_{j+1/2} \neq \tilde{g}^n_{j+1/2}$ if and only if

$$(8.9)\quad |\tilde{g}^n_{j+1/2}| = \min(|\tilde{g}^n_{j-1/2}|, |\tilde{g}^n_{j+1/2}|, |\tilde{g}^n_{j+3/2}|).$$

It follows from (7.5) that

$$|\hat{g}^n_{j+1/2}| = \min\left(|\hat{g}^n_{j-1/2}|, |\hat{g}^n_{j+1/2}|, |\hat{g}^n_{j+3/2}|\right)$$

and therefore (8.5) follows from (8.8) and (8.9)

$$|\tilde{g}^{n+1}_{j+1/2}| \le |\tilde{g}^n_{j+1/2}| = \min\left(|\tilde{g}^n_{j-1/2}|, |\tilde{g}^n_{j+1/2}|, |\tilde{g}^n_{j+3/2}|\right)$$

This completes the proof of Lemma 8.1.

$C_{\Delta,n}$ is a monotonicity preserving operator, and since w^0_j is a monotone function of j, so is w^n_j. Furthermore, since w^0_j is a strictly monotone function of j in $J_{R,0} - j - J_{L,0} \le j \le J_{R,0}$, we get from Lemma 8.1 that

Corollary 8.2. w^n_j is a strictly monotone function of j in $J_{L,n} \le j \le J_{R,n}$.

The proof follows immediately from (8.5) and (7.3).

Lemma 8.3. $J_{R,\infty} - J_{L,\infty} \le 2$.

Proof by negation: It follows from the definition of $J_{L,\infty}$ and $J_{R,\infty}$ that

(8.10) $\tilde{g}^n_{j+1/2} \to 0$ as $n \to 0$ for $j \le J_{L,\infty}-1$ and $j \ge J_{R,\infty}$,

(8.11) $|\tilde{g}^n_{J_{L,\infty}+1/2}| \ge \eta_1 > 0, \quad |\tilde{g}^n_{J_{R,\infty}-1/2}| \ge \eta_2 > 0$, $n \ge 0$.

Now we assume that $J_{R,\infty} - J_{L,\infty} \ge 3$ and arrive at a contradiction. Let $\eta_3 = \min_{J_{L,\infty}+1 \le j \le J_{R,\infty}-2} |\tilde{g}^0_{j+1/2}|$, and let $\eta = \min(\eta_1, \eta_2, \eta_3)$. Hence

(8.12) $$\min_{J_{L,\infty} \le j \le J_{R,\infty}-1} |\tilde{g}^0_{j+1/2}| \ge \eta .$$

It follows from (8.5) and (8.12) that for $J_{L,\infty}+1 \le j \le J_{R,\infty}-2$

$$|\tilde{g}^1_{j+1/2}| \ge \min\left(|\tilde{g}^0_{j-1/2}|, |\tilde{g}^0_{j+1/2}|, |\tilde{g}^0_{j+3/2}|\right) \ge \eta$$

From (8.11) we get that $|\tilde{g}^1_{J_{L,\infty}+1/2}| \ge \eta$ and $|g^1_{J_{R,\infty}-1/2}| \ge \eta$ and therefore $\min_{J_{L,\infty} \le j \le J_{R,\infty}-1} |\tilde{g}^1_{j+1/2}| \ge \eta > 0$.
By induction,

$$\min_{J_{L,\infty} \le j \le J_{R,\infty}-1} |\tilde{g}^n_{j+1/2}| \ge \eta > 0 , \qquad n \ge 0 . \tag{8.13}$$

Multiplying (7.10a) by sgn = sign $(v_R - v_L)$, we get

$$\begin{aligned} \text{sgn. } w^{n+1}_{J_{L,\infty}+1} = \text{sgn. } w^n_{J_{L,\infty}+1} &- \hat{\lambda}\, \delta_{J_{L,\infty}+3/2} \\ &\left(|g^n_{J_{L,\infty}+2}| - |g^n_{J_{L,\infty}+1}|\right) \\ &- \hat{\lambda}(1-\delta_{J_{L,\infty}+1/2}) \left(|g^n_{J_{L,\infty}+1}| - |g^n_{J_{L,\infty}}|\right) \end{aligned} \tag{8.14}$$

It follows from (8.13) that

$$\begin{aligned} |g^n_{J_{L,\infty}+1}| &= \min\left(|\hat{g}^n_{J_{L,\infty}+1/2}|, |\hat{g}^n_{J_{L,\infty}+3/2}|\right), \\ &\ge \min\left(|\tilde{g}^n_{J_{L,\infty}+1/2}|, |\tilde{g}^n_{J_{L,\infty}+3/2}|\right) \ge \eta \end{aligned} \tag{8.15a}$$

and similarly

$$|g^n_{J_{L,\infty}+2}| \ge \eta > 0 . \tag{8.15b}$$

On the other hand, we get from (8.10) that

$$|g^n_{J_{L,\infty}}| = \min\left(|\hat{g}^n_{J_{L,\infty}-1/2}|, |\hat{g}^n_{J_{L,\infty}+1/2}|\right) \to 0 \text{ as } n \to \infty .$$

Therefore $\exists N$ such that

$$|g^n_{J_{L,\infty}}| \le \eta/2 \quad \text{for all} \quad n \ge N . \tag{8.15c}$$

Substituting (8.15) into (8.14), we get

$$\text{sgn. } w^{n+1}_{J_{L,\infty}+1} \leq \text{sgn. } w^{n}_{J_{L,\infty}+1} - \hat{\lambda}\,\eta/2 \quad \text{for all} \quad n \geq N .$$

Thus for all $k > 0$, we get

$$\text{sgn. } w^{N+k}_{J_{L,\infty}+1} \leq \text{sgn. } w^{N}_{J_{L,\infty}+1} - \frac{k}{2}\,\hat{\lambda}\,\eta \tag{8.16}$$

Subtracting sgn. v_L from both sides of (8.16) we get

$$\begin{aligned} |w^{N+k}_{J_{L,\infty}+1} - v_L| &= \text{sgn } (w^{N+k}_{J_{L,\infty}+1} - v_L) \\ &\leq |w^{N}_{J_{L,\infty}+1} - v_L| - \frac{1}{2}\,k\,\hat{\lambda}\,\eta . \end{aligned} \tag{8.17}$$

Thus by taking k to be sufficiently large in (8.17), $|w^{N+k}_{J_{L,\infty}+1} - v_L|$ can be made negative, which is a contradiction.

Lemma 8.4. Let w^{∞} be defined as in (8.4), then

$$C_{\Delta} w^{\infty} = w^{\infty} . \tag{8.18}$$

Proof: $\tilde{g}_{j+1/2} = 0$ for all $j \neq J_{\infty}-1, J_{\infty}$. Hence

$$\begin{aligned} \text{sgn. } g_j &= \min \left(|\hat{g}_{j-1/2}|, |\hat{g}_{j+1/2}|\right) \\ &= \min \left(|\tilde{g}_{j-1/2}|, |\tilde{g}_{j+1/2}|\right) = 0 , \quad j \neq J_{\infty} \end{aligned}$$

and $\delta_{J_{\infty}-1/2} = 0$, $\delta_{J_{\infty}+1/2} = 1$. Therefore it follows immediately that $(C_{\Delta} w^{\infty})_j = w^{\infty}_j$ for $j \neq J_{\infty}$. At $j = J_{\infty}$ $(C_{\Delta} w^{\infty})_{J_{\infty}} = w^{\infty}_{J_{\infty}} - \hat{\lambda}(0 - g_{J_{\infty}}) - \hat{\lambda}(1-0)(g_{J_{\infty}} - 0) = w^{\infty}_{J_{\infty}}$ which completes the proof of Lemma 8.4.

The main theorem of this section is an immediate consequence of Lemmas 18.3 and 18.4.

Theorem 8.5. For all $0 < \hat{\lambda} \leq 1$, $w^n = (\prod_{k=0}^{n-1} C_{\Delta,k})w^0$ converges as $n \to \infty$ to a stationary shock-like solution.

We remark that the exact shape of w, i.e. J_∞ and in (8.4), is determined solely by w^0 (independent of $\hat{\lambda}$). C_Δ is a finite-difference operation in conservation form which is consistent with the flux g. But $g_j^n = 0$ for $j \leq J_{L,0}$ and $j \geq J_{R,0}$ and therefore

$$\sum_{j=J_{L,0}}^{J_{R,0}} w_j^{n+1} = \sum_{j=J_{L,0}}^{J_{R,0}} w_j^n \quad \text{for all} \quad n \geq 0 .$$

Hence

$$\sum_{j=J_{L,0}}^{J_{R,0}} w_j^\infty = \sum_{j=J_{L,0}}^{J_{R,0}} w_j^0 \tag{8.19}$$

Thus J_∞ = integer and $0 \leq \alpha \leq 1$ are uniquely determined by

$$(J_\infty - J_{L,0} + 1 - \alpha)v_L + (J_{R,0} - J_\infty + \alpha)v_R = \sum_{j=J_{L,0}}^{J_{R,0}} w_j^0 \tag{8.20}$$

We finish this section with the following theorem concerning a special property of the convergence for $\hat{\lambda} = 1$.

Theorem 8.6. Let C_Δ be the operator in (10.7) with $\hat{\lambda} = 1$. Then $w^n = (\prod_{k=0}^{n-1} C_{\Delta,k})w^0$ converges to w^∞ in a finite number of time steps.

Proof: Denote $J_{L,N} = J_0$ and $J_{R,N} = J_1$. To prove the theorem we have to show that if $J_1 - J_0 \geq 3$, then $\exists k = k_N$

such that

$$J_{R,N+k} - J_{L,N+k} \le J_1 - J_0 - 1 . \tag{8.21}$$

If for some $n_0 \ge N$, $|\tilde{g}_{J_0+1/2}| \le |\tilde{g}_{J_0+5/2}|$, then it follows from (7.5) that $|\tilde{g}_{J_0+1/2}| = |\hat{g}_{J_0+1/2}| \le |\hat{g}_{J_0+3/2}| \le |\hat{g}_{J_0+5/2}|$. From (7.7) it follows that $g_{J_0} = 0$, $g_{J_0+1} = \tilde{g}_{J_0+1/2}$ and $g_{J_0+2} = \hat{g}_{J_0+3/2}$. Substituting this into (7.10) with $\hat{\lambda} = 1$, we get

$$w^{n_0+1}_{J_0+1} = w^{n_0}_{J_0+1} - g^{n_0}_{J_0+1/2} = w^{n_0}_{J_0+1} - (w^{n_0}_{J_0+1} - w^{n_0}_{J_0}) = w^{n_0}_{J_0} = v_L .$$

Therefore $J_{L,n_0} = J_{L,N} + 1$ and (8.21) follows. Similarly if for some $n_0 \ge N$, $|\tilde{g}_{J_1-1/2}| \le |\tilde{g}_{J_1-5/2}|$ then $J_{R,n_0} = J_{R,N} - 1$ which implies (8.21).

We shall prove the theorem by negating the possibility that

$$|\tilde{g}_{J_0+1/2}| > |\tilde{g}_{J_0-5/2}| \text{ and } |\tilde{g}_{J_1-1/2}| > |\tilde{g}_{J_1-5/2}| \quad \text{for all } n \ge N. \tag{8.22}$$

As in Lemma 18.3, it follows from (8.5) that

$$\min_{J_0+1\le j\le J_1-1} |\tilde{g}^{n+1}_{j+1/2}| \ge \min_{J_0\le j\le J_1-1} |\tilde{g}^{n}_{j+1/2}| .$$

If (8.22) is true for all $n \ge N$, then

$$\min_{J_0\le j\le J_1-1} |\tilde{g}^{n+1}_{j+1/2}| = \min_{J_0+1\le j\le J_1-2} |\tilde{g}^{n+1}_{j+1/2}| \le \min_{J_0\le j\le J_1-1} |\tilde{g}^{n}_{j+1/2}|$$

$$\le \cdots \le \min_{j_0\le j\le J_1-1} |\tilde{g}^{N}_{j+1/2}| = \eta > 0 \quad \text{for all } n \ge N .$$

In particular

(8.23) $\quad |\tilde{g}^n_{J_0+5/2}| \geq \eta > 0 \quad \text{for all} \quad n \geq N \ .$

But if $|\tilde{g}_{J_0+1/2}| > |\tilde{g}_{J_0+5/2}|$ then from (7.5) $|\hat{g}_{J_0+3/2}| \geq |\hat{g}_{J_0+5/2}|$ and therefore $|g_{J_0+1}| \geq |g_{J_0+2}|$ (7.7). Since $g_{J_0} = 0$, we get by substituting these relations into (7.10) that $\text{sgn. } w^{n+1}_{J_0+1} = \text{sgn. } w^n_{J_0+1} - |g_{J_0+2}|$
$= \text{sgn} - \min\left(|\hat{g}_{J_0+3/2}|, |\hat{g}_{J_0+5/2}|\right) \leq \text{sgn. } w^n_{J_0+1} - |\tilde{g}_{J_0+5/2}|$
$\leq \text{sgn. } w^n_{J_0+1} - \eta \ .$

The last inequality follows from (8.23). Since this inequality holds for all $n \geq N$ we get for all $k \geq 0$ that

(8.24) $\quad \text{sgn. } w^{N+k}_{J_0+1} \leq \text{sgn. } w^N_{J_0+1} - k\,\eta$

Subtracting $\text{sgn. } v_L$ from both sides of (8.24) we get

(8.25) $\quad |w^{N+k}_{J_0+1} - v_L| \leq |w^N_{J_0+1} - v_L| - k\,\eta.$

Thus by taking k to be sufficiently large in (8.25), $|w^{N+k}_{J_0+1} - v_L|$ can be made negative, which is a contradiction.

9. Existence of Steady Progressing Solutions of L_Δ.

In this section we shall discuss existence of steady progressing profiles of the corrective type ACM

$$L_\Delta = (C_\Delta)^K L_1 , \qquad K \geq 1 \tag{9.1}$$

corresponding to a Riemann problem, where the initial discontinuity is either a shock or a contact discontinuity. L_1 is a first order accurate approximation to the original equation (1.3a) and C_Δ is the artificial compressor (7.10). First we shall discuss how large K should be in order that L_Δ will produce a steady progressing solution, and then present numerical verification.

Let L_1 be a centered finite difference scheme using 2J+1 points; then at most 2J additional values are diffused per one application of L_1. From Theorem 8.6 we get that if $\hat{\lambda} = 1$ in (7.10), then we can choose a sufficiently large finite K in (9.1), such that the diffused discontinuity will be compressed to its minimal representation in the finite mesh, i.e. 1-2 cells depending on its location. As was already observed by Von Neumann and Richtmyer [24], the numerical gradient $g_{j+1/2}$ (7.3) has typically a single extremum located at the discontinuity. Thus, using the notations of Theorem 8.6, $|\tilde{g}_{J_0+1/2}| \leq |\tilde{g}_{J_0+5/2}|$ and $|\tilde{g}_{J_1-1/2}| \leq |\tilde{g}_{J_1-5/2}|$ is typical to regions of discontinuity; these imply $J_{L,n+k} = J_{L,n}+k$ and $J_{R,n+k} = J_{R,n}-k$ and therefore $K = J$ is sufficient to prevent smearing (see proof

of Theorem 8.6).

In the following we shall consider the LF scheme (3.5) which has J = 1. We shall show that K = 1 is adequate in the sense that L_Δ (9.1) produces steady progressing profiles which have almost maximal resolution. We use the term steady progressing in the sense of (3.18), that is, we say that a finite difference scheme possesses a steady progressing solution travelling at speed S if for any rational number $\eta = \lambda S$, where $\lambda = \Delta t/\Delta x$ is in the range of the stability condition, there exists an integer P_η such that ηP_η = integer and for all j

$$v_j^{n+P_\eta} = v_{j-\eta P_\eta}^n \tag{9.2}$$

We consider first the contact discontinuity problem

$$u_t + u_x = 0 \tag{9.3a}$$

$$u(x,0) = S_0(x) = \begin{cases} 1 , & x < 0 \\ 0 , & x \geq 0 \end{cases} \tag{9.3b}$$

Let $w^n(\lambda,\hat{\lambda})$ be the numerical solution to (9.3) obtained by (9.1) where L_1 is the Lax-Friedrichs scheme

$$v_j^{n+1} = (L_1 v^n)_j = \frac{1}{2}(v_{j+1}^n + v_{j-1}^n) - \frac{1}{2}\lambda (v_{j+1}^n - v_{j-1}^n) , \tag{9.4}$$

$0 \leq \hat{\lambda} \leq 1$ is defined in (7.10). We have found that for all those rational numbers λ which were tested, $w^n(\lambda,1)$ becomes a steady progressing solution after a relatively small number of time steps. In Table II we show the

periodicity of $w^n(\frac{1}{2},1)$, which is $P_{1/2} = 4$, and compare this steady progressing solution with the transition layer obtained by the Lax-Friedrichs scheme (9.4) after 100 Δt, without the artificial compressor C_Δ. Observe that the two cells around the exact location of the contact discontinuity contain 85% of the jump on the average; the minimum is 76% and the maximum is 94%. Note that after 100 Δt, 68% of the jump in the LF solution already occupies about 20 cells. Additional computations verified the $\sqrt{n}$ behavior of the smearing, and indeed after 400 Δt, 68% of the jump occupies 40 cells, which is $\sqrt{4}$ times that of 100 Δt. In Table III we give the periodic solutions of $w^n(\lambda,1)$ for $\lambda = 0,1/4,1/2,1$. Because of this and additional computations with other rational values of λ, we make the conjecture that $w^n(\lambda,1)$ is periodic for all rational λ, that is, L_Δ possesses a steady progressing solution. Observe that $w(\lambda,1)$ does not depend strongly on λ, and the profiles are rather similar. As λ increases the dissipation term decreases (see (3.8b)) and the resolution is slightly improved.

Next we consider the shock problem

$$u_t + (\tfrac{1}{2} u^2)_x = 0 \tag{9.5a}$$

$$u(x,0) = S_0(x) \ . \tag{9.5b}$$

$S_0(x)$ is the unit step function defined in (9.3b). Again we denote by $w^n(\lambda,\hat{\lambda})$ the numerical solution to (9.5) obtained by (9.2) where L_1 is the Lax-Friedrichs scheme ($\hat{\lambda}$ as before);

$$(9.6)\quad v_j^{n+1} = (L_1 v^n)_j = \frac{1}{2}(v_{j+1}^n + v_{j-1}^n) - \frac{1}{4}\lambda[(v_{j+1}^n)^2 - (v_{j-1}^n)^2]$$

In this case the LF scheme itself possesses a steady progressing solution (see Jennings' theorem 3.2), hence we expect $w^n(\lambda,\hat{\lambda})$ to have a periodic dependence on n for all rational numbers $0 \leq \lambda \leq 1$ and all real numbers $0 \leq \hat{\lambda} \leq 1$. This was found to be true in all the cases that were tested. In Table IV we give the periodic solutions of $w(\lambda,\hat{\lambda})$ for $\hat{\lambda} = 1$ with $\lambda = 0, 1/4, 1/2, 3/4, 1$; for comparison's sake, $w(\frac{1}{2},\frac{1}{2})$ and $w(\frac{1}{2},0)$ are also given; the latter is the LF scheme without the artificial compressor. Observe that the resolution of the shock in $w(\lambda,1)$ is better than that of a contact discontiuity; the two cells around the exact location of the shock contain 96% of the jump, compared to 85% of the jump for the contact discontinuity. We can see from Table IV that the dependence of the resolution on the amount of the artificial compressor is very strong: 96% of the jump in $w(\frac{1}{2},\frac{1}{2})$ occupy 7 cells compared to 2 cells in $w(\frac{1}{2},1)$.

10. Limitations of the ACM.

The ACM was designed to prevent smearing of contact discontinuities and to improve resolution of shocks. This is done by artificially making the characteristic field $\frac{dx}{dt} = a(u)$ more convergent by the amount $\gamma_{j+1/2}$ (7.13b), i.e. $\frac{d\hat{x}}{dt} = a(u) + \gamma(u)$. In order to make the method automatic we apply it everywhere and construct it so that it adds terms of $O((\Delta x)^2)$ in all regions of smoothness. In this section we shall describe the nature of the error caused by the ACM in smooth linear problems, and discuss the applicability of the method to rarefaction waves.

First we discuss smooth linear problems and we consider an initial data $u_0(x)$ with bounded support. One application of the LF scheme L_1 (4.5) shifts the solution with approximately the right speed, but decreases the values of the maxima, increases the values of the minima and enlarges the support. It is easy to show that as $n \to \infty$, $(L_1)^n u_0(x)$ approaches zero. On the other hand, an application of the artificial compressor C_Δ to $u_0(x)$ preserves the location of the extrema and their values, and sharpens the transition between successive extrema. As $n \to \infty$, $(C_\Delta)^n u_0(x)$ approaches a square-wave structure, the constant states of which are the extrema of $u_0(x)$ (see Theorem 8.5).

We state without proof that an application of the combined operator $L_\Delta = C_\Delta \cdot L_1$ with $\hat{\lambda} = 1$ to $u_0(x)$, decreases its maxima, increases its minima (as L_1 alone) but sharpens the transition between successive extrema. As $n \to \infty$, $(L_\Delta)^n u_0(x)$ approaches a square-wave structure; each constant state corresponds to an extremum of $u_0(x)$. A constant state corresponding to a maximum of $u_0(x)$ has a smaller value, and the one corresponding to a minimum has a larger value.

The above analysis shows that neither the LF scheme, nor the modified $L_\Delta = C_\Delta \cdot L_1$ method gives satisfactory results for long term computation in this case. This is in spite of the fact that both methods converge to the exact solution as $\Delta x \to 0$ with $t \leq T$.

The application of the ACM that we have in mind is to nonlinear conservation laws, where the only linearity is in the form of a contact discontinuity; there the property of overcompression in regions of linearity will be an advantage: If the contact discontinuity is formed during the computation, it will be compressed to its steady progressing form -- e.g. see the Mach-reflection calculation in Section 14. To check how well a smeared contact discontinuity is restored we performed the following numerical experiment: We applied the LF scheme (3.4) with $\lambda = 0.8$ to (9.3b) for 20 time steps; since no artificial compression was applied the discontinuity was smeared. Then we continued

the computation with $L_\Delta = C_\Delta \cdot L_1$ with $\tilde{\lambda} = 1$ and $\lambda = 0.8$ as before. In another 20 time steps the discontinuity was completely restored (i.e. it approached the steady progressing solution of Table III).

Next we consider the application of L_Δ to nonlinear problem. The characteristics (1.6) are either convergent or divergent; in the first case, this region will develop into a shock and the worst that can happen is that it will be formed a little earlier than it should -- this does not affect the accuracy in later times. A divergent characteristic field represents a rarefaction. If the characteristic field is almost degenerate, i.e. the characteristics are divergent but almost parallel, then the ACM should not be applied since it might prevent the development of the rarefaction and even compress it into an artificial contact discontinuity.

If the characteristic field diverges sufficiently strongly, this can be prevented: $|\gamma_{j+1/2}| \approx \frac{(\Delta x)|u_{xx}|}{|u_x|}$, but in a rarefaction wave $u(x,t) = h(\frac{x}{t})$ (see (1.18)) and therefore $u_{xx} = \frac{1}{t^2} h''(\xi)$, $u_x = \frac{1}{t} h'(\xi)$ and $|\gamma_j| = \frac{\Delta x}{t} \frac{h''(\xi)}{h'(\xi)}$; $\xi = x/t$. If we start the computation without applying any artificial compression for, say, 10 time steps or more, $|\gamma_{j+1/2}|$ which decreases like $1/t$ will become sufficiently smaller than $|a(u)|$, the characteristic speed of the original problem. When we apply the ACM to this data, it does not interfere with the development

of the rarefaction. As was mentioned earlier, the ACM is capable of recompressing a smeared contact discontinuity, and therefore no harm is done by this procedure.

However, it is adivsable not to apply the ACM to rarefaction waves whenever possible. Rarefaction wave regions can be excluded by the following technique: The discontinuity region is the steepest and therefore the maximum of the absolute value of the numerical gradient is obtained there. We propose to compute the maximal absolute value of the numerical gradient at each time step, and to apply C_Δ only to those monotone regions where the numerical gradient is comparable in size to this maximal value.

We remark that another possible way to exclude a rarefaction wave is to use a technique similar to that of Rosenbluth -- see [22], p. 313.

11. Extension of the ACM to Multi-Dimensional Problems.

The ACM can be easily extended to conservation laws in any number K of space dimensions:

$$u_t + \sum_{k=1}^{K} \frac{\partial}{\partial x_k} f_k(u) = 0 \tag{11.1}$$

To simplify notation we take K = 2;

$$u_t + p(u)_x + q(u)_y = 0 \tag{11.2}$$

The most direct extension is to use schemes which are combinations of one-dimensional operators L_{x_k}, and to replace each of them by $\hat{L}_{x_k} = C_{x_k} L_{x_k}$ where C_{x_k} is a one-dimensional artificial compressor in the x_k-direction. Let us consider for example the Lax-Friedrichs approximation to (11.2).

$$v^{n+1}_{i,j} = (L_1 v^n)_{i,j} = \tfrac{1}{4}\left(v^n_{i+1,j}+v^n_{i-1,j}+v^n_{i,j+1}+v^n_{i,j-1}\right) - \tfrac{1}{2}\lambda_x[p(v^n_{i+1,j})-p(v^n_{i-1,j})] - \tfrac{1}{2}\lambda_y[q(v^n_{i,j+1})-q(v^n_{i,j-1})] \tag{11.3}$$

where $v^n_{i,j} = v(i\Delta x, j\Delta y, n\Delta t)$, $\lambda_x = \Delta t/\Delta x$ and $\lambda y = \Delta t/\Delta y$. L_1 can be written as

$$L_1 = \tfrac{1}{2}(L_{1x}+L_{1y}) \tag{11.4a}$$

where L_{1x} is the one-dimensional LF opeator in the x-direction approximating

$$u_t + p(u)_x = 0$$

$$w_i = (L_{1x}v)_i = \tfrac{1}{2}(v_{i+1}+v_{i-1}) - \lambda_x[p(v_{i+1})-p(v_{i-1})] , \tag{11.4b}$$

which is stable under the condition $\lambda_x \max_u |\frac{dp}{du}| \leq \frac{1}{2}$; L_{1y} is the one-dimensional LF operator in the y-direction approximating

$$u_t + q(u)_y = 0$$

(11.4c) $w_j = (L_{1y}v)_j = \frac{1}{2}(v_{j+1}+v_{j-1}) - \lambda_y[q(v_{j+1})-q(v_{j-1})]$

which is stable under the condition $\lambda_y \max_u |\frac{dq}{du}| \leq \frac{1}{2}$.
Thus L_1 is stable under the condition

(11.4d) $\max \{\lambda_x \max_u |\frac{dp}{du}|, \lambda_y \max_u |\frac{dq}{du}|\} \leq \frac{1}{2}$.

The ACM can be extended by replacing L_1 in (11.4a) by

(11.5) $\hat{L}_1 = \frac{1}{2}(\hat{L}_{1x}+\hat{L}_{1y}) = \frac{1}{2}(C_x L_{1x}+C_y L_{1y})$

where C_x and C_y are one-dimensional artificial compressions in the x and y direction, respectively. The stability of $\hat{L}_1$ and its compressiveness follow directly from those of the one-dimensional operators.

We remark that this approach is not limited to first order accurate schemes. In general it is possible to construct a scheme approximating (11.1) to any degree of accuracy by using a suitable combination of one-dimensional operators having the same degree of accuracy. Furthermore, it was shown by G. Strang [23] (see also [8]) that an optimal stability condition can be obtained e.g. one can construct a second order L_2 by taking the following combination of the second order accurate one-dimensional operators L_{2x} and L_{2y}

(11.6) $$L_2 = \frac{1}{2}\left(L_{2x}L_{2y} + L_{2y}L_{2x}\right)$$

The ACM can be extended by replacing L_{2x} and L_{2y} by $\hat{L}_{2x} = C_x L_{2x}$ and $\hat{L}_{2y} = C_y L_{2y}$, where C_x and C_y are the one-dimensional artificial compressors such that C_x, C_x, $C_y = I + O((\Delta x)^3)$ in regions of smoothness -- see a forthcoming report.

Although Strang-type schemes are very efficient compared to other standard schems (see [5]), it is desirable to extend the ACM in such a way that any finite difference scheme for several space dimensions could be used. To do so, we follow the same general pattern of the one-dimensional case. We consider the initial value problem (11.2) with initial data of two constant states u_- and u_+ separated by the straight line

(11.7a) $$(x-x_0)\sin\alpha - (y-y_0)\cos\alpha = 0$$

the normal to which is

(11.7b) $$\nu = (\sin\alpha, -\cos\alpha)$$

Let S be the quantity defined by (1.20), i.e.

(11.7c) $$S = \frac{[p(u_+)-p(u_-)]\sin\alpha - [q(u_+) - q(u_-)]\cos\alpha}{u_+ - u_-}$$

We assume that (u_-, u_+, S) is a contact discontinuity or a shock i.e. inequality (1.26) holds for all u between u_- and u_+.

Let $G(u) = (g^x(u), g^y(u))$ be a C^1 vector function of u with the following properties:

(11.8a) $$G(u_-) = G(u_+) = 0$$

and for all u between u_- and u_+

(11.8b) $$G(u) \parallel \nu$$

(11.8c) $$G(u)\cdot\nu/(u_+-u_-) > 0 .$$

Then G(u) is called an <u>artificial compression function</u> (ACF).

Formula (11.8b) implies that

(11.9) $$g^x(u) = g_0(u) \sin\alpha, \quad g^y(u) = -g_0(u)\cdot\cos\alpha$$

It follows from (11.8a) and (11.8c) that

$$g_0(u_-) = g_0(u_+) = 0$$

and that for all u between u_- and u_+

$$g_0(u)/(u_--u_+) > 0$$

Thus any ACF for the 2-dimensional problem is given by (11.9), where $g_0(u)$ is a one-dimensional ACF satisfying (5.2) and (5.4) with respect to the discontinuity (u_-,u_+,S). As in the one-dimensional case, if the discontinuity of u when regarded as the solution of (11.1) is a contact discontinuity on a shock, then it is a shock when regarded as the soluiton of

(11.10) $$u_t + [p(u)+g^x(u)]_x + [q(u)+g^y(u)]_y = 0$$

We propose to approximate (11.10) by a corrective-type ACM $\hat{L}$:

(11.11) $$v^{n+1} = \hat{L}v^n = C\cdot Lv^n$$

L is any stable first order accurate approximation to (11.1), and C is a finite difference approximation to

(11.12a) $$u_t + g^x(u)_x + g^y(u)_y = 0$$

This can be accomplished by taking C to be

$$C = \frac{1}{2}(C_x + C_y)$$

where C_x is a one-dimensional stable approximation to

(11.12b) $$u_t + g^x(u)_x = 0$$

and C_y is a one-dimensional stable approximation to

(11.12c) $$u_t + g^y(u)_y = 0$$

The stability of $\hat{L}$ (11.11) follows immediately from the stability of L and of the one-dimensional operators C_x and C_y.

Next we show that the Δ-ACM can be extended to the multi-dimensional case in the same way. We consider a numerical approximation to the initial value problem (11.1) with the initial data (11.7). In the numerical solution the discontinuity will be represented as a transition layer in the normal direction; in this layer we regard the numerical solution as a function $u^N(z)$ of the normal coordinate $z = (x-x_0)\cos\alpha + (y-y_0)\sin\alpha$. The transition in the x-direction $u(x,y_0)$ through the point (x_0,y_0) is given by $u(x,y_0) = u^N((x-x_0)\sin\alpha)$; the transition in the y-direction $u(x_0,y)$ through this point is $u(x_0,y) = u^N(-(y-y_0)\cos\alpha)$. Hence for $\Delta x = \Delta y = h$ we get

(11.13a) $u(x+\Delta x,y_0) - u(x,y_0) \approx \frac{d}{dz} u^N(z)\ h \sin \alpha$

(11.13b) $u(x_0,y+\Delta y) - u(x_0,y) \approx - \frac{d}{dz} u^N(z)\ h \cos \alpha$.

Let g^x be the Δ-ACF in the x-direction obtained by applying (7.3), (7.5) and (7.7) to $u(x,y_0)$; let g^y be the Δ-ACF in the y-direction obtained by applying these operations to $u(x_0,y)$. From (7.3) and (11.13) we get that

$$\tilde{g}^x \approx \frac{d}{dz} u^N(z)\ h \sin \alpha\ , \quad \tilde{g}^y \approx - \frac{d}{dz} u^N(z)\ h \cos \alpha$$

therefore

(11.14a) $$\tilde{g}^x/\sin \alpha \approx - \tilde{g}^y/\cos \alpha$$

Let F be the flattening operation in (7.5). It is easily verified that for any grid function v_j and any real number α, $F\cdot(\alpha v) = \alpha F\cdot v$; the same is true for the operation (7.7). Therefore it follows from (11.14a) that

(11.14b) $$g^x/\sin \alpha \approx -g^y/\cos \alpha$$

Hence (11.9) is satisfied and (g^x,g^y) is a 2-dimensional ACF.

Let $C_{\Delta x}$ be the one-dimensional Δ-artificial compressor defined by (7.10) with respect to g^x , and let $C_{\Delta y}$ be the one defined with respect to g^y. We extend the Δ-ACM to the multi-dimensional case by replacing L in (11.11) by L_Δ

(11.1-) $$L_\Delta = C_\Delta\ L\ , \qquad C_\Delta = \frac{1}{2}\ (C_{\Delta x} + C_{\Delta y})\ .$$

The smearing in the multi-dimensional case occurs along the normal direction. The Δ-ACM reduces to the one-dimensional case along this direction; as in the

one-dimensional case we expect the Δ-ACM to produce steady progressing profiles whenever (u_-, u_+, S) is a shock or a contact discontinuity. Indeed this was verified in all our computer experiments. For example let us consider the contact discontinuity problem

(11.16a) $$u_t + u_x + u_y = 0$$

with the initial data of two constant states $u_- = 1$ and $u_+ = 0$ separated by the straight line (11.7a) with $\alpha = 135°$. We used L_Δ in (11.15) with L being the LF-scheme (11.3) with $\lambda_x = \lambda_y = 0.4$, and $C_\Delta = \frac{1}{2}(C_{\Delta x} + C_{\Delta y})$ where $\hat{\lambda}_x = \hat{\lambda}_y = 1$ (7.10a). The numerical solution approached a steady progressing profile. $v(x, y_0)$, the transition in the x-direction and $v(x_0, y)$ the transition in the y-direction of this profile were the same:

1.00, 0.98, 0.85, 0.18, 0.

We remark that the scheme

$$\hat{L}_1 = \frac{1}{2}\,(C_{\Delta x} L_{1x} + C_{\Delta y} L_{1y})$$

where L_{1x} and L_{1y} are the one-dimensional LF operators (11.4), produces similar results.

V. Hyperbolic System of Conservation Laws -- Weak Solutions and Their Numerical Computation

12. Weak Solutions of a Hyperbolic System of Conservation Laws.

A system of conservation laws is an equation of the form

$$u_t + f(u)_x = 0 \tag{12.1}$$

where u is a vector of m components and f(u) a vector valued function of u. The system of conservation laws is called hyperbolic if $A(u) = \text{grad}_u f$ has real and distinct eigenvalues $\lambda_1(u),\ldots,\lambda_m(u)$ for all values of the argument u; we assume $\lambda_k(u)$ are arranged in increasing order. We denote by $r_k(u)$ the right eigenvector of A(u) corresponding to the eigenvalue $\lambda_k(u)$.

The theory of hyperbolic systems of conservation laws has these features in common with the theory of a single conservation law: i) If the equations are genuinely nonlinear, that is, for all $1 \leq k \leq m$ and all values of u

$$r_k(u)\cdot\text{grad}_u\lambda_k(u) \neq 0 \tag{12.2}$$

then the initial value problem cannot be solved for all time unless we admit weak solutions; these are solutions of the integral version of the conservation laws (12.1).
ii) A piecewise smooth weak solution satisfies the Rankine-Hugoniot (RH) condition

$$f(u_R) - f(u_L) = S(u_R - u_L) \tag{12.3}$$

across the lines of discontinuity.

iii) Weak solutions are not uniquely determined by their initial values; an additional criterion is needed for selecting the physically relevant weak solutions. Motivated by considerations of the theory of free boundary value problems, Lax [14] suggested the following shock condition: A jump discontinuity in a weak solution is called a shock if the total number of characteristics drawn with respect to the state on the left u_L in the positive t-direction, which stay to the left of the discontinuity curve, and those drawn with respect to the state on the right u_R, which stay to the right of the discontinuity curve, is m-1. The analytical-expression for this requirement is: for some index k, $1 \leq k \leq m$, the inequalities

$$\lambda_{k-1}(u_L) < S < \lambda_k(u_L)$$
$$\lambda_k(u_R) < S < \lambda_{k+1}(u_R) \tag{12.4}$$

hold. k will be called the index of the shock; there are m different kinds of shocks.

Lax [14] proved that any given state can be connected to a one-parameter family of nearby states on the right through a k-shock provided that the k-th family of characteristics is genuinely nonlinear.

Foy [6] related the shock defined by (12.4) to the weak solutions which are selected by the viscosity method (Section 2). He proved that if two nearby states can be connected through a shock, then they can be connected through a viscous profile, i.e. a steady progressing solution

of

$$u_t + f(u)_x = \varepsilon u_{xx} \tag{12.5}$$

Next we discuss the case where the k-th characteristic field is linearly degenerate, i.e. for all values of u

$$r_k(u) \cdot \text{grad}_u \lambda_k(u) \equiv 0 \ . \tag{12.6}$$

A function $v(u)$ is a k-Riemann invariant if for all values of u

$$r_k(u) \cdot \text{grad}_u \, v \equiv 0 \ . \tag{12.7}$$

Using this terminology, the fact that the k-th field is linearly degenerate is expressed by saying that λ_k is a k-Riemann invariant. The significance of condition (12.6) is that it guarantees the existence of travelling wave solutions, i.e. solutions of the form $u(x-ct) \equiv u(y)$. Substituting this form of solution in (12.1) we arrive at the equation

$$- c \frac{d}{dy} u + A \frac{d}{dy} u = 0 \ .$$

Consequently $\frac{d}{dy} u = \alpha \, r(u(y))$, $c = \lambda(u(y))$ and

$$\frac{dc}{dy} = \text{grad}_u \, c \cdot \frac{du}{dy} = \alpha \, \text{grad}_u \, \lambda \cdot r(u) \ .$$

This condition (12.6) implies that c = constant, and a travelling wave solution exists.

In this case it is possible that a solution of (12.1) will have a "linear discontinuity" which should be admitted as physically relevant. The following theorem and definition is due to Lax [14].

Theorem 12.1. If two nearby states u_L and u_R have the same k-Riemann invariants with respect to a degenerate field, then the RH conditions (12.3) are satisfied with

$$S = \lambda_k(u_L) = \lambda_k(u_R) \tag{12.8}$$

Such a discontinuity is called a contact discontinuity.

Proof: (Lax [14]): We have to show that

$$S(u_R - u_L) = f(u_R) - f(u_L)$$

Imagine u_R and u_L connected by a differentiable one-parameter family of states $u(\theta)$ for which all Riemann invariants are constant. This is certainly possible if u_R and u_L are nearby states. Since λ_k is one of the Riemann invariants, $\lambda_k(u(\theta)) \equiv S$ for all θ. For all other k-Riemann invariants $v(u)$

$$0 = \frac{d}{d\theta} v(u(\theta)) = \operatorname{grad}_u v \cdot \frac{d}{d\theta} u(\theta) \ . \tag{12.9}$$

There are exactly m-1 independent k-Riemann invariants, the gradients of which span the orthogonal complement of r_k. $\frac{d}{d\theta} u(\theta)$, being orthogonal to all k-Riemann invariants (12.9), is proportional to $r_k(u(\theta))$; therefore

$$A \frac{d}{d\theta} u = S \frac{d}{d\theta} u \ . \tag{12.10}$$

Integrating this we obtain the RH relation.

Corollary. If two nearby states u_L and u_R are connected by a contact discontinuity, then there exists a differentiable one-parameter family of states $u(\theta)$ connecting u_L and u_R such that

(12.11) $\quad g(u(\theta)) \equiv g(u_L) = g(u_R) = C$

where $g(u) = f(u) - Su$.

Proof: Take $u(\theta)$ to be the differentiable one-parameter family of the previous theorem; they by (12.10)

$$\frac{d}{d\theta} g(u(\theta)) = (A\text{-}SI) \frac{du}{d\theta} = 0 .$$

Thus $g(u(\theta))$ = constant; (12.11) follows from the RH relation $g(u_L) = g(u_R)$ (we denote their common value by C).

Lax's definition of a contact discontinuity is restricted to nearby states. In view of the corollary we suggest removing this restriction by defining a contact discontinuity in the following way.

Definition 12.2. Let u_L and u_R be two states satisfying the RH relation with some S. This discontinuity will be called a contact discontinuity if and only if there exists a differentiable one-parameter family of states $u(\theta)$ connecting u_L and u_R along which (12.11) holds.

Remark. This one parameter family consists of states $u(\theta)$ which satisfy the RH condition relative to u_L and u_R with speed S.

Definition 12.2 generalizes Lax's definition in the following sense.

Theorem 12.3. Let u_L, u_R and $u(\theta)$ be as in Definition 12.2. Then

(a) u_L and u_R have the same k-Riemann invariants;

(b) $S = \lambda_k(u_L) = \lambda_k(u_R)$;

(c) the k-th characteristic field is linearly degenerate for all values of $u(\theta)$.

Proof: Differentiating (12.11) we get

$$0 = \frac{d}{d\theta}\,[g(u(\theta)) - C] = (A(u(\theta)) - SI)\,\frac{du}{d\theta}$$

Hence: $S = \lambda_k(u(\theta)) \equiv \text{const.}$, $\frac{du}{d\theta} = \alpha\, r_k(u(\theta))$. The first identity implies $S = \lambda_k(u_L) = \lambda_k(u_R)$. Differentiating it we get

$$0 = \frac{d}{d\theta}\,\lambda_k(u(\theta)) = \text{grad}_u\lambda_k \cdot \frac{du}{d\theta} = \alpha\ \text{grad}_u\lambda_k\ r_k$$

i.e. the k-th characteristic field is linearly degenerate for all $u(\theta)$. In addition, if v is any of the k-Riemann invariants, we get from (12.7) that

$$\frac{dv}{d\theta} = \text{grad}_u v \cdot \frac{du}{d\theta} = \alpha\ \text{grad}_u v \cdot r_k = 0$$

i.e. the k-Riemann invariants are the same for both sides as well as constant along $u(\theta)$.

To exemplify these notions let us consider the Eulerian equations of a gas in conservation form

(12.12a) $$W_t + F(W)_x = 0$$

(12.12b) $$W = \begin{pmatrix} \rho \\ m \\ E \end{pmatrix}, \quad F(W) = \left[u \begin{pmatrix} \rho \\ m \\ E \end{pmatrix} + \begin{pmatrix} 0 \\ p \\ pu \end{pmatrix} \right] = uW + \begin{pmatrix} 0 \\ p \\ pu \end{pmatrix}$$

ρ, m, and E are the mass, momentum and total energy per unit volume, respectively; $u = m/\rho$ is the velocity, and p is related to W by an equation of state. $A(W) = \text{grad}_W F$

has three distinct eigenvalues u, $u \pm c$ where c is the sound speed. The two Riemann invariants corresponding to the eigenvalue u are u and p. W_L and W_R are connected by a contact discontinuity corresponding to this eigenvalue if there exists a differntiable one-parameter family of states, $W(\theta)$ connecting W_L and W_R along which $u(\theta)$=const., $p(\theta)$ = const.

If we assume the equation of state to be that of a polytropic gas, i.e.

$$p = (\gamma-1)(E - \frac{1}{2}\rho u^2) \tag{12.13}$$

where γ is the polytropic constant, we get that if W_L and W_R satisfy the RH relation with $S = u_L = u_R$ and $p_L = p_R$, then $g(W(\theta)) \equiv g(W_L) = g(W_R)$, along

$$W(\theta) = W_L + \theta(W_R - W_L) , \qquad 0 \leq \theta \leq 1. \tag{12.14}$$

13. ACM for Hyperbolic Systems of Conservation Laws.

In the case of the single conservation law, it was found that Oleinik's condition E (1.10) selects the physically relevant weak solutions. We have expressed this condition in the form (1.15)

$$(13.1) \quad [g(u)-C]/(u_R-u_L) > 0 , \quad C = g(u_L) = g(u_R)$$

where g, the so-called compression function, is defined by (1.9)

$$(13.2) \quad g(u) = f(u) - Su$$

Numerical calculations and analysis show that in the discrete approximation to solutions of (1.3), where a shock is represented as a transition over a narrow region, the width of the transition region is inversely proportional to the size of g(u)-C: To produce a narrower transition region we have to artificially raise this quantity. For contact discontinuity, where $g(u)-C \equiv 0$, an artificial compression term is absolutely necessary.

We turn now to discuss discrete approximations to solutions of the system of conservation laws (12.1). We consider initial data of two constant states u_L and u_R which satisfies the RH condition (12.3) for some S, and denote

$$(13.3) \quad g_\ell(u) = f_\ell(u) - Su_\ell , \qquad 1 \le \ell \le m$$

We shall show that if the LF scheme (3.5) produces a steady

progressing profile V_j^n corresponding to this initial data, then the numerical gradient of the ℓ-th component is proportional to $g_\ell(V)-C$; consequenty the width of the transition in the ℓ-th component is inversely proportional to $|g_\ell(V)-C|$. For the sake of simplicity we consider a stationary profile, i.e. S = 0. In this case g(u) = f(u) and $C = f(u_L) = f(u_R)$. A stationary profile V_j of the LF scheme satisfies the following difference equation:

$$V_j = \frac{1}{2}(V_{j+1}+V_{j-1}) - \frac{1}{2}\lambda[f(V_{j+1})-f(V_{j-1})] \tag{13.4}$$

By rearranging (13.4) we get

$$V_{j+1}-V_j-\lambda[f(V_{j+1})+f(V_j)] = V_j-V_{j-1}-\lambda[f(V_j)+f(V_{j-1})]$$

$$= V_{j-1}-V_{j-2}-\lambda[f(V_{j-1})+f(V_{j-2})] = \ldots = -2\lambda C .$$

Denoting $f_{j+1/2} = \frac{1}{2}[f(V_{j+1})+f(V_j)]$ we get for all j

$$V_{j+1}-V_j = 2\lambda(f_{j+1/2}-C) . \tag{13.5}$$

Next we turn to discuss computation of contact discontinuities. The LF scheme can be written as

$$v_j^{n+1} = \tfrac{1}{2}(v_{j+1}^n+v_{j-1}^n) - \tfrac{1}{2}\lambda S(v_{j+1}^n-v_{j-1}^n) - \tfrac{1}{2}\lambda[g(v_{j+1}^n)-g(v_{j-1}^n)] , \tag{13.6a}$$

where g is defined by (13.3). Computations of contact discontinuities of a polytropic gas (12.12)-(12.13) by the LF scheme, show clearly that the velocity u and the pressure p in the transition region are constant to a large number of digits. It follows from the analysis

presented at the end of Section 12, that in this case $g(v_j^n) \equiv C$ for all j and n. Hence (13.6a) for a polytropic gas is practically

$$(13.6b) \quad v_j^{n+1} = \frac{1}{2}\,(v_{j+1}^n + v_{j-1}^n) - \frac{1}{2}\,S(v_{j+1}^n - v_{j-1}^n)$$

That is, the computation by the LF scheme of contact discontinuities of a polytropic gas is practically the computation of m <u>independent</u> linear discontinuities. Indeed our calculations, as well as earlier calculations by Lax [13], show that the width of the transition is proportional to $\sqrt{n}$ as is expected in this case from considerations of a single linear discontinuity (see Section 4).

In general, g(v) in the computation of a contact discontinuity is by its definition (12.2) not only continuous but also close to a constant. Therefore differences of g(v) are usually negligible relative to differences of v itself which is discontinuous. Thus the computation of a contact discontinuity in general has the nature of (13.6b). This, and the fact that in calculations of shocks the width of the transition in the ℓ-th component is inversely proportional to the size of $|g_\ell(v)-C|$, suggest that a componentwise extension of the ACM will be in order; that is, to use an ACF $\hat{g}(u)$,

$$(13.7a) \qquad \hat{g}(u) = \begin{pmatrix} \hat{g}_1(u) \\ \vdots \\ \hat{g}_m(u) \end{pmatrix}$$

such that

$$\hat{g}_\ell(u) = \hat{g}_\ell(u_\ell) \tag{13.7b}$$

where $\hat{g}_\ell(u_\ell)$ is a scalar ACF with respect to the ℓ-th component.

Our main goal is to prevent the smearing of a contact discontinuity, which was shown to behave approximately like a linear discontinuity. Let us denote by w_j^n the scalar numerical solution by (13.6b) to the initial data of the two constant states $W_L = 0$ and $W_R = 1$. We expect $(v_\ell)_j^n$ the ℓ-th component of v_j^n, to be approximately

$$(v_\ell)_j^n \approx u_{L,\ell} + w_j^n(u_{R,\ell}-u_{L,\ell}) , \qquad 1 \le \ell \le m \tag{13.8}$$

in the locality of the contact discontinuity. Let $g_0(w)$ be a scalar ACF defined by (5.2) and (5.4) for $w_L = 0$ and $w_R = 1$, and let $\hat{w}_j^n$ be the numerical solution by an ACM using the ACF $g_0(w)$. We propose to use $\hat{g}_\ell$ of the form

$$\hat{g}_\ell(u_\ell) = g_0\left(\frac{u_\ell - u_{L,\ell}}{u_{R,\ell}-u_{L,\ell}}\right)(u_{R,\ell}-u_{L,\ell}), \quad 1 \le \ell \le m \tag{13.9}$$

This way we expect $\hat{v}_j^n$, the numerical solution by a componentwise extension of the ACM using (13.9) to be approximately

$$\hat{v}_j^n \approx u_L + \hat{w}_j^n(u_R-u_L) . \tag{13.10}$$

Next we show that this componentwise extension is indeed an ACM.

Theorem 13.1. Let (u_L,u_R,S) be either a shock or a contact discontinuity with respect to $u_t+f(u)_x = 0$; and let $\hat{g}(u)$ be defined by (13.9). Then (u_L,u_R,S) is a shock with respect to

$$u_t + [f(u) + \hat{g}(u)]_x = 0 . \tag{13.11}$$

Proof: It follows from the assumption that (u_L,u_R,S) is either a shock or a contact discontinuity that

$$\lambda_k(u_L) \geq S \geq \lambda_k(u_R) \tag{13.12}$$

where $\lambda_k(u)$ are the eigenvalues of $A(u) = \text{grad}_u f$. Denote by $\gamma_0(w) = dg_0(w)/dw$; the assumption that $g_0(w)$ is an ACF for $w_L = 0$ and $w_R = 1$ that (5.4)

$$\gamma_0(0) > 0 \quad \text{and} \quad \gamma_0(1) < 0 . \tag{13.13}$$

Let $\hat{A}(u) = \text{grad}_u(f+\hat{g})$ and denote its eigenvalues by $\hat{\lambda}_k(u)$. It follows from (13.9) that

$$\hat{A}(u_L) = A(u_L) + \gamma_0(0)I , \quad \hat{A}(u_R) = A(u_R) + \gamma_0(1)I$$

where I is the unit matrix. Therefore

$$\hat{\lambda}_k(u_L) = \lambda_k(u_L) + \gamma_0(0), \quad \hat{\lambda}_k(u_R) = \lambda_k(u_R) + \gamma_0(1) . \tag{13.14}$$

It follows from (13.12)-(13.14) that

$$\hat{\lambda}_k(u_L) > \lambda_k(u_L) \geq S \geq \lambda_k(u_R) > \hat{\lambda}_k(u_R)$$

and therefore (u_L,u_R,S) is a shock for (13.11) by definition 12.4.

Remark. In order not to change the characteristic field outside the region of the discontinuity in an explicit ACM, we have to choose $g_0(w)$ such that $\gamma_0(1) = \gamma_0(1) = 0$ but $\gamma_0(\varepsilon) > 0$ and $\gamma_0(1-\varepsilon) < 0$ for all $\varepsilon > 0$ sufficiently small. In the case of a single equation this gives rise to discontinuities for which $a(u_L) = S = a(u_R)$ but for all $\varepsilon > 0$ sufficiently small,

$$a(u_L+\varepsilon(u_R-u_L)) > S > a(u_R-\varepsilon(u_R-u_L))$$

Oleinik's condition E (1.15) admits such discontinuities as shocks; likewise it seems that the strict inequality at the endpoints in Lax's shock condition for systems (12.4) is too severe and may be relaxed to

$$\lambda_k(u_L+\varepsilon(u_R-u_L)) > S > \lambda_k(u_R-\varepsilon(u_R-u_L))$$

for all $\varepsilon > 0$ sufficiently small.

Our next numerical example supports this proposal. We consider a contact discontinuity problem for a polytropic gas (12.12)-(12.13) with $\gamma = 1.4$:

$$\text{(13.15a)}\quad w(x,0) = \begin{cases} w_L, & x \le 0 \\ & \\ w_R, & x > 0 \end{cases}, \quad w_L = \begin{pmatrix} \rho_L \\ m_L \\ E_L \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 10 \end{pmatrix}, \quad w_R = \begin{pmatrix} \rho_R \\ m_R \\ E_R \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 8 \end{pmatrix}$$

We solve it by the LF approximation to

$$\text{(13.15b)}\quad W_t + [F(W)+\alpha\hat{g}(W)]_x = 0, \quad \hat{g}(W) = \begin{pmatrix} (\rho-\rho_L)^2(\rho-\rho_R)^2/(\rho_R-\rho_L)^3 \\ (m-m_L)^2(m-m_R)^2/(m_R-m_L)^3 \\ (E-E_L)^2(E-E_R)^2/(E_R-E_L)^3 \end{pmatrix}$$

$\lambda = \Delta t/\Delta x$ is taken to be 95% of the maximal permissible value. After 400 time steps we find that for $\alpha = 0$ ($T = 9.9$), i.e. no artificial compression, the width of the transition is already 54 cells; for $\alpha > 0$ there is a steady progressing profile. For $\alpha = 10$ ($T = 7.5$) the width of the transition is 5 cells, and for $\alpha = 20$ ($T = 6.0$) the width is 3 cells. T is the time after 400 Δt. Although $\gamma_0(0) = \gamma_0(1) = 0$ (13.13), the ACM produces "shock-like" profiles. In all three cases the transition is monotone in each of the components, and u and p in the transition region are constant to a large number of digits.

A corrective type ACM for a system of conservation laws has the form

$$\hat{v}^{n+1} = L\, v^n \tag{13.16a}$$

where L is some standard finite difference scheme, and

$$v^{n+1} = C\, \hat{v}^{n+1}, \quad C = \begin{pmatrix} C_1 & & \\ & \ddots & \\ & & C_m \end{pmatrix} \tag{13.16b}$$

where C_ℓ is a finite difference approximation to the single equation

$$w_t + (u_{R,\ell} - u_{L,\ell})\, g_0\left\{\frac{w - u_{L,\ell}}{u_{R,\ell} - u_{L,\ell}}\right\}_x = 0 . \tag{13.17}$$

If C_ℓ, $1 \le \ell \le m$ and L have L_2-stable first variation (see [22]), so has $\hat{L} = C\, L$. Since $(u_{L,\ell}, u_{R,\ell})$ is a stationary shock of (13.17), a good choice for C_ℓ is the one-sided scheme (3.11).

Next we turn to discuss a componentwise extension of the Δ-ACM. From (13.8) we get that

$$(13.18)\quad (v_\ell)^n_{j+1} - (v_\ell)^n_j \approx (w^n_{j+1}-w^n_j)(u_{R,\ell}-u_{L,\ell})$$

Let us denote by $g^n_{\Delta,0}$ the ACF assigned to w^n_j by the operations (7.3), (7.5) and (7.7); let us denote by $g^n_{\Delta,\ell}$, $1 \le \ell \le m$ the ACF assigned to $(v_\ell)^n_j$ by the same operations. It follows from (13.18) and the nature of the operations (7.3), (7.5) and (7.7) that

$$(13.19)\qquad g^n_{\Delta,\ell} \approx (u_{R,\ell}-u_{L,\ell})\, g^n_{\Delta,0}\,, \qquad 1 \le \ell \le m\,.$$

Thus a componentwise extension of the Δ-ACM falls into the same category of extensions discussed before and the same analysis applies.

$$(13.20a)\qquad L_\Delta = C_\Delta \cdot L$$

$$(13.20b)\qquad C_\Delta = \begin{pmatrix} C_{\Delta,1} & & \\ & \ddots & \\ & & C_{\Delta,m} \end{pmatrix}$$

where $C_{\Delta,\ell}$ is the artificial compressor defined by (7.10) with respect to $(v_\ell)_j$, the ℓ-th component of v_j. For the sake of uniformity, $C_{\Delta,\ell}$ are all taken with $\hat{\lambda} = 1$.

Finally we turn to test (13.20) numerically. Again we consider the contact discontinuity problem for a polytropic gas (13.15a). We take L to be the LF approximation to (12.12)-(12.13) with stability condition,

(13.21) $\lambda \max_W(|u|+c) \leq 1 , \quad c^2 = \gamma p/\rho .$

We find that indeed L_Δ produces steady progressing profiles. The results of the calculation with $\lambda = 0.25$ which is about 96% of the maximal value of (13.21) are given in Table V. As can be seen from the table, the resolution is good: 85% of the jump is contained in 2 cells. It is interesting to note that u and p are constant to a considerable number of digits; this is an a posteriori justification to the assumptions (13.8) and (13.10)

14. Numerical Results.

In this section we describe some nontrivial numerical experiments carried out with the scheme

$$L_{\Delta} = C_{\Delta} \cdot L_1 \tag{14.1}$$

for the Riemann problem in one-dimension and the ordinary oblique reflection and the Mach reflection in two dimensions; C_{Δ} is the Δ-artificial compressor (13.20b) and L_1 is the LF scheme.

Frigure 3 shows the results of two calculations for the Riemann problem (12.12)-(12.13) with the initial data

$$(14.2) \quad \begin{cases} \rho(x) = 0.445 \\ u(x) = 0.698 \\ p(x) = 3.528 \end{cases}, \quad x < 0 \; ; \quad \begin{cases} \rho(x) = 0.5 \\ u(x) = 0. \\ p(x) = 0.571 \end{cases}, \quad x \geq 0$$

$\gamma = 1.4$. Results of calculations of the Riemann problem with the same initial data but in Lagrange coordinates are presented in [4] and [18]. A shock wave develops and propagates to the right followed by a rarefaction wave which propagates to the left. A contact discontinuity which moves with the local fluid velocity lies between the shock and the rarefaction wave.

Figures 3a and 3c show the results of the calculation by the LF scheme ($\lambda/\lambda_{max} = 0.95$) without any artificial compression ($\hat{\lambda} = 0$) after 100 Δt and 200 Δt. The density in the region between the contact discontinuity and the

shock is somewhat lower than its exact value, and after 200 Δt, 80% of the jump of the contact discontinuity occupy 34 cells.

Figures 3b and 3d show the results of the calculation by the ACM (14.1) with $\lambda/\lambda_{max} = 0.95$ and $\hat{\lambda} = 1$ after 100 Δt and 200 Δt. C_Δ was not applied to the rarefaction wave; it was excluded by the technique which was suggested at the end of Section 10. The density in the region between the contact discontinuity and the shock oscillates slightly around its exact value. Observe that both the shock and the contact discontinuity have the same spread of 2 cells. The pressure and the velocity have some overshoot in the transition across the contact discontinuity which do not exist in the actually computed quanties ρ, m and E. This overshoot is due to the fact that the artificial compression is only approximately uniform with respect to all components (13.19) and it decreases in time.

Next we discuss calculations of ordinary oblique reflection and Mach reflection. We consider a shock tube with a wedge of angle α. We neglect viscosity and other dissipative mechanisms and we assume a perfect-gas equation of state.

$$p = (\gamma-1)\ [E - \tfrac{1}{2}\ (m^2+n^2)/\rho]\ . \tag{14.3}$$

We use the following conservation form of these equations.

$$W_t + F(W)_x + G(W)y = 0 \tag{14.4a}$$

$$(14.4b)\quad W = \begin{pmatrix} \rho \\ m \\ n \\ E \end{pmatrix}, \quad F(W) = \begin{pmatrix} m \\ \frac{3-\gamma}{2}\frac{m^2}{\rho} + (\gamma-1)\left(E - \frac{1}{2}\frac{n^2}{\rho}\right) \\ \frac{mn}{\rho} \\ \gamma\frac{Em}{\rho} + \frac{\gamma-1}{2}\frac{m}{\rho^2}(m^2+n^2) \end{pmatrix},$$

$$G(W) = \begin{pmatrix} n \\ \frac{nm}{\rho} \\ \frac{3-\gamma}{2}\frac{n^2}{\rho} + (\gamma-1)\left(E - \frac{1}{2}\frac{m^2}{\rho}\right) \\ \gamma\frac{En}{\rho} + \frac{\gamma-1}{2}\frac{n}{\rho^2}(m^2+n^2) \end{pmatrix}$$

Here ρ is the mass per unit volume, the density, and $m = \rho u$, and $n = \rho v$ is the momenta per unit volume in the x and y directions; E is the total energy per unit volume.

In this problem an incident shock changes the flow direction through the angle α so that it is parallel to the wedge; a seconeary or reflecting shock from the upper wall changes the flow by an equal amount so that it is again parallel to the upper wall -- see Figure 4a. Considering the problem as being time-dependent allows the flow to approach the steady state shock reflection solution when arbitrary initial data are prescribed. In our calculation we took the wedge to have an angle $\alpha = \tan^{-1}(\frac{1}{5})$ and we prescribed an initial condition of impulsive flow, i.e. $W(x,y,0) = W_0$ everywhere; the state W_0 is given by $\rho_0 = 1$, $m_0 = 1$, $v_0 = 0$ and $M_0 = 2.2$ with $\gamma = 1.4$ in (14.3). The exact incident shock angle θ_1 is 37.2° and the exact reflected shock angle θ_2 is 46.6° (see [2]). The boundary conditions used are similar to those of [2].

Figure 4b shows the density field of the steady-state obtained by the LF scheme with $\lambda/\lambda_{max} = 0.95$ and without any artificial compression, $\hat{\lambda} = 0$. Figure 4c shows the density field of the steady-state obtained by the ACM L_Δ (14.1); L_1 is the LF scheme with $\lambda/\lambda_{max} = 0.95$ as before and $C_\Delta = \frac{1}{2}(C_{\Delta x} + C_{\Delta y})$ with $\hat{\lambda} = 0.9$. The solution by the ACM is a good approximation to the exact solution: The angles θ_1 and θ_2 are c rrect and the density between the incident and the r flected shock oscillates slightly around the exact constant value which is 1.567. The shocks are rather narrow; the spread becomes larger near the boundaries -- this is probably because of the first order accuracy of the boundary conditions.

When $\alpha = \tan^{-1}(\frac{1}{5})$ and W_0 such that $\rho_0 = m_0 = 1$, $v_0 = 0$, $M_0 = 1.6$ and $\gamma = 1.2$, oblique reflection can no longer occur. Approximate theories and experimental evidence indicate that the exact solution has the structure of the Mach reflection illustrated in Figure 5 (see [2] and the references cited there; see also [21]). The line separating regions C and D represents a contact discontinuity or a slip line; all other lines represent shocks.

Figures 6a-6d show the actual values of ρ, m, n and the Mach number M obtained after 1000 time steps by the LF scheme with $\lambda/\lambda_{max} = 0.95$. The computed discontinuities are heavily smeared and the configuration of Figure 5 is hardly recognizable; the smearing of the contact discontinuity is complete.

Figures 7a-7d show the values of the same quantities obtained after 1000 time stpes by the ACM (14.1). L_1 is again the LF scheme with $\lambda/\lambda_{max} = 0.95$ and $C_\Delta = \frac{1}{2}(C_{\Delta x} + C_{\Delta y})$ with $\hat{\lambda} = 1$. The resolution of the shocks and the contact discontinuity is very good. The incident shock angle θ_1 is about 50°, and the angle of the reflected shock θ_2 is approximately 67°.

S. Burstein [2] computed the same problem by the two-dimensional Lax-Wendroff scheme. This scheme produces an incident shock angle $\theta_1 = 52.3°$ and a reflected shock angle $\theta_2 = 72.4°$. It was found that the slipline emerging from the triple point could not be numerically resolved in this calculation.

We remark that the calculations of the Mach reflection presented here, as well as those presented in [2] are only temporarily steady. The flow near the right boundary is subsonic, a fact that allows disturbances generated by the extrapolating procedure at the right boundary, to propagate upstream. When the calculation is continued beyond 1000 time steps, the noise from the right boundary begins to dominate the numerical results.

Acknowledgement.

I am more than thankful to Peter Lax, my advisor, for his encouragement and patience. Working with him was very instructive as well as pleasant. I would like to thank Sam Burstein for many interesting and helpful discussions.

References

1. J. P. Boris, and D. L. Book, "Flux Corrected Transport I. SHASTA, A Fluid Transport Algorithm that Works," Jour. Comp. Phys., Vol. , 1973, pp.
2. S. Z. Burstein, "Numerical Methods in Multidimensional Shocked Flows," AIAA Jour., Vol. 2, 1964, pp. 211-217.
3. S. Z. Burstein, "Finite Difference Calculations for Hydrodynamic Flows Containing Discontinuities," Jour. Comp. Phys., Vol. 1, 1966, pp. 198-222.
4. S. Z. Burstein and A. A. Mirin, "Third Order Difference Methods for Hyperbolic Equations," Jour. Comp. Phys., Vol. 5, 1970, pp. 547-571.
5. B. Eilon, D. Gottlieb and G. Zwas, "Numerical Stabilizer and computing time for second-order accurate schemes," Jour. of Comp. Phys., Vol. 9, 1972, pp. 387-397.
6. R. L. Foy, "Steady State Solutions of Hyperbolic Systems of Conservation Laws with Viscosity Terms," Comm. Pure Appl. Math., Vol. 17, 1964, pp. 177-188.
7. S. K. Godunov, "Finite Difference Methods for Numerical Computation of Discontinuous Solutions of Equations of Fluid Dynamics," Mat. Sb., Vol. 47, 1959, pp. 271-295.
8. D. Gottlieb, "Strang type difference schemes for multidimensional problems," SIAM Jour. Num. Anal., Vol. 9, 1972, pp. 650-661.
9. A. Harten, and G. Zwas, Switched Numerical Shumin Filters for Shock Calculations," Jour. of Eng. Math., Vol. 6, 1972, pp. 207-216.

10. A. Harten and G. Zwas, "Self-Adjusting Hybrid Schemes for Shock Computations," Jour. of Comp. Phys., Vol. 9, 1972, pp. 568-583.

11. G. Jennings, "Discrete Travelling Waves," to appear in Comm. Pure Appl. Math.

12. B. Keyfitz, "Solutions with Shocks, an Example of an L_1-Contractive Semigroup," Comm. Pure Appl. Math., Vol. 24, 1971, pp. 125-132.

13. P. D. Lax, "Weak Solutions of Nonlinear Hyperbolic Equations and Their Numerical Computation," Comm. Pure Appl. Math., Vol. 7, 1954, pp. 159-193.

14. P. D. Lax, "Hyperbolic Systems of Conservation Laws II," Comm. Pure Appl. Math., Vol. 10, 1957, pp. 537-566.

15. P. D. Lax, "On the Stability of Difference Approximations to Solutions of Hyperbolic Equations with Variable Coefficients," Comm. Pure Appl. Math., Vol. 14, 1961, pp. 497-520.

16. P. D. Lax, "Nonlinear Partial Difference Equations and Computing," SIAM Review, Vol. 11, 1969, pp. 7-19.

17. P. D. Lax, "Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves," Regional Conference Series in Applied Mathematics, No. 11, 1973, Published by SIAM.

18. P. D. Lax and B. Wendroff, "Systems of Conservation Laws," Comm. Pure Appl. Math., Vol. 13, 1960, pp. 217-237.

19. O. A. Oleinik, "Discontinuous Solutions of Nonlinear Differential Equations," Uspekhi Mat. Nauk (N.S.,

Vol. 12, 1957, pp. 3-73 (Amer. Math. Soc. Transl. Ser. 2, 26, pp. 95-172).

20. S. A. Orszay and L. W. Jayne, "Local Errors of Difference Approximations to Hyperbolic Equations," Jour. of Comp. Phys., Vol. 14, 1974, pp. 93-103.
21. R. D. Richtmyer, "Progress Report on the Mach Reflection Calculation," NYO 9764, Courant Institute of Mathematical Sciences, New York Univ. (Sept. 1961).
22. R. D. Richtmyer and K. W. Morton, "Finite Difference Methods for Initial Value Problems," Interscience-Wiley, New York, 1967.
23. G. Strang, "On the Construction and Comparison of Difference Schemes," SIAM Jour. Num. Anal., Vol. 5, 1968, pp. 506-517.
24. J. Von Neumann and R. D. Richtmyer, "A Method for the Numerical Calculation of Hydrodynamic Shocks," Jour. Appl. Phys., Vol. 21, 1950, pp. 232-237.
25. C. B. Vrugdenhill, "On the Effect of Artificial Viscosity Methods in Calculating Shocks," Jour. Eng. Math., Vol. 3, 1969, pp. 285-287.
26. G. Zwas and J. Roseman, "The Effect of Nonlinear Transformations on the Computation of Weak Solutions," Jour. Comp. Phys., Vol. 12, 1973, pp. 179-186.

Appendix A. Viscous Approximation to Numerical Solutions of Shocks.

We have shown in Section 3 that solutions of the LF scheme

$$(A1)\quad v_j^{n+1} = (L_1 v^n)_j = \frac{1}{2}(v_{j+1}^n + v_{j-1}^n) - \frac{1}{2}\lambda[f(v_{j+1}^n) - f(v_{j-1}^n)]$$

can be approximated by solutions of the corresponding viscous equation

$$(A2)\quad w_t + f(w)_x = \Delta t[\beta(w,\lambda)w_x]_x\ ,\quad \beta(w,\lambda) = \frac{1}{2}\left[\frac{1}{\lambda^2} - \left(\frac{df}{dw}\right)^2\right]$$

provided that $w(x,t)$ is sufficiently smooth. In Section 4 we have shown that in the linear case $f = S\,u$, $S = \text{const.}$, this statement remains true even if the initial data is discontinuous. We remark that the artificial viscosity methods are based on the idea that this approximation is valid even for shock solutions, i.e. that addition of terms which in shock regions have the nature of the truncation error of the LF scheme, makes the numerical solution behave similarly to a solution of a corresponding viscous equation. In this appendix we shall present some more numerical evidence to support the validity of this approximation for shock solutions.

First we discuss a recent result by Zwas and Roseman [26], which is exemplified by the following: Consider the initial value problem

$$(A3)\quad u_t + \left(\frac{u^2}{2}\right)_x = 0$$

with the discontinuous initial data

$$(A4) \qquad u(x,0) = \phi(x) = \begin{cases} 1, & x < 0 \\ 0 & x \geq 0 \end{cases}$$

Solving this problem by the LF scheme (A1) we get a steady progressing solution which propagates with the correct speed S = 1/2. Now, under the transformation $v = u^2$, (A3) becomes

$$(A5) \qquad v_t + (\tfrac{2}{3} v^{3/2})_x = 0$$

Solving (A5) with the same initial data by the LF scheme we get a steady progressing solution which propgates with a speed 2/3. Assuming that in the original problem the LF scheme really solves

$$(A6) \qquad u_t + (\tfrac{1}{2} u^2)_x = \tfrac{1}{2} \Delta t[(\frac{1}{\lambda^2} - u^2)u_x]_x$$

Zwas and Roseman suggested modifying the LF scheme for the transformed problem (A5) such that it will approximate the transformed version of (A6)

$$(A7) \qquad v_t + \tfrac{2}{3} v^{3/2}{}_x = \tfrac{1}{2} \Delta t\{[(\frac{1}{\lambda^2} - v)v_x]_x + (1 - \frac{1}{\lambda^2 v})v_x^2\}$$

Doing so they got the same weak solution of the original problem, i.e. a speed of propagation which is equal to 1/2. This indicates that the LF scheme really approximates solutions of (A2), which are smooth even for discontinuous initial data, and therefore remain unchanged under transformation.

Finally we compare profiles of the LF approximation to (A4) (i.e. (A1) with $f = \frac{1}{2} u^2$) to profiles of the corresponding viscous equation (A6). For the sake of simplicity we consider a stationary shock

$$u(x,0) = \begin{cases} 1, & x < 0 \\ -1, & x \geq 0 \end{cases}$$

The profiles of (A6) are given by $z = x/\Delta x$, $\lambda = \Delta t/\Delta x$,

$$z = -\lambda u - \frac{1-\lambda^2}{2\lambda} \ln \left(\frac{1+u}{1-u}\right) \tag{A8}$$

(see 2.9a), where $z = x/\Delta x$ and $\lambda = \Delta t/\Delta x$.

In Table VI profiles of (A1) and (A8) are compared for $\lambda = 0.2, 0.4, 0.6, 0.8$. As can be seen from this table the agreement is very good and the numerical profile is much closer to the viscous profile than to the original step function.

Appendix B. Higher Order Linear Approximations to a Contact Discontinuity.

We consider a R-th order accurate linear finite difference scheme $v_j^{n+1} = (L_R v^n)_j$; approximating solutions of the linear equation

$$w_t + S\, w_x = 0 , \quad S = \text{const.} \tag{B1}$$

L_R approximates more closely solutions of

$$w_t + S\, w_x = \frac{(\Delta x)^{R+1}}{\Delta t}\, \beta_R(\lambda)\, \frac{\partial^{R+1} w}{\partial x^{R+1}} \tag{B2}$$

provided $w(x,t)$ is sufficiently smooth; $\beta_R(\lambda)$ can be found by the same expansion procedure described in Section 3.

First we shall show that this statement is still true even if the initial data is discontinuous; then we shall use the solution of (B2) to get the leading term in the time dependence of the width of the numerical solution.

When considered in the system of coordinates which moves with the discontinuity (B2) reduces to

$$w_t = \frac{(\Delta x)^{R+1}}{\Delta t}\, \beta_R(\lambda)\, \frac{\partial^{R+1} w}{\partial y^{R+1}} \tag{B3}$$

where y measures the deviation from the center of the discontinuity. If the discontinuous initial data is composed of two constant states, then we get by a similarity argument that

$$w(y,t) = F\left(\frac{y}{\Delta x \left[\frac{\beta_R(\lambda)}{\Delta t}\, t \right]^{1/(R+1)}} \right) \equiv F(z) . \tag{B4}$$

Formula (B2) was so constructed that

$$(B5)\quad w(x,t+\Delta t) - L_R w(x,t) = \text{const.}\ (\Delta t)^{R+2}\ \frac{\partial^{R+2} w}{\partial x^{R+2}} \equiv E(x,t)\ .$$

Using (B4) we get that for $t \geq \delta > 0$ and $\lambda = \Delta t/\Delta x = \text{const.}$

$$E(x,t) = \text{const.}\ (\Delta t)^{\frac{R+2}{R+1}}\ \frac{d^{R+2}F(z)}{dz^{R+2}}\ ,$$

z is defined in (B4).

Consequently $\|E\|_{L_2}^2 = \int_{-\infty}^{\infty} |E(x,t)|^2\ dx$

$$= \text{const.}\ (\Delta t)^{2(R+2)/(R+1)} \int_{-\infty}^{\infty} \left|\frac{d^{R+2}F(z)}{dz^{R+2}}\right|^2\ dy$$

$$= \text{const.}\ (\Delta t)^{2(R+2)/(R+1)}\ \frac{\Delta x}{(\Delta t)^{1/(R+1)}} \int_{-\infty}^{\infty} \left|\frac{d^{R+2}F(z)}{dz^{R+2}}\right|^2 dz$$

$$= \text{const.}\ (\Delta t)^{(3R+4)/(R+1)}\ .\quad \text{Hence}$$

$$(B6)\quad \|w(x,t+\Delta t) - L_R w(x,t)\|_{L_2} = \text{const.}\ (\Delta t)^{\frac{3R+4}{2(R+1)}}\ .$$

By using the same method as Lax's equivalence theorem (see [22]), we get that for all $0 < \delta \leq t \leq T$,

$$(B7)\quad \|w(x,t) - v(x,t)\|_{L_2} = \text{const.}\ (\Delta t)^{\frac{3R+4}{2(R+1)} - 1} = \text{const.}\,(\Delta t)^{\frac{R+2}{2(R+1)}}.$$

$v(x,t)$ is the numerical solution by L_R.

On the other hand, the L_2 difference between the numerical solution $v(x,t)$ and the exact solution $u(x,t)$ for this discontinuous initial data is given by (see [20] and the references cited there)

$$(B8)\quad \|v(x,t) - u(x,t)\|_{L_2} = \text{const.}\ (\Delta t)^{\frac{R}{2(R+1)}}\ .$$

Comparing (B7) with (B8) we see that indeed $v(x,t)$ approximates solutions of (B2) to a higher degree of accuracy than it approximates solutions of the original equation (B1). In addition, the leading term in $v(x,t)-u(x,t)$ is contained in $w(x,t)-u(x,t)$. From (B4) we get that $w(x,t) = \text{const.}$ along $y/\Delta x = \text{const.}\ [\frac{\beta(\lambda)}{\Delta t}\ t]^{\frac{1}{R+1}}$, and therefore the spread w_R is proportional to

$$w_T = \text{const.}\ [\beta(\lambda)\ n]^{\frac{1}{R+1}} \tag{B9}$$

Vrengdenhill [25] gives a comparison between solutions of the LF and the LW scheme with solutions of the corresponding equations (B2); a good agreement was found. If R = even, then the (R+1)-th derivative represents a dispersive term. In this case solutions of (B2) have oscillations which are in phase with those of the numerical solutions, but do not have the same amplitude. Addition of the dissipative (R+2)-th derivative to (B2) takes care of the discrepancy in amplitude and the resulting solution is in excellent agreement with the numerical solution. (A similar phenomenon occurs in the nonlinear case, see [26].)

Appendix C. Antidiffusion Methods.

It was shown in Section 3 that a first order scheme L_1 can be represented by the equation (see (3.24))

$$w_t + f(w)_x = \Delta t[\beta(w,\lambda)w_x]_x \tag{C1}$$

and that the width of the shock produced by this scheme can be approximated by (see (3.27))

$$W(u_-,u_+) = \lambda \int_{u_-}^{u_+} \frac{(w,\lambda)}{g(w)-C}\, dw \tag{C2}$$

Thus, W is proportional to $\beta(u,\lambda)$, the coefficient of the diffusion term.

This observation motivates the antidiffusion method, which is a stable modification of a diffusive finite difference scheme; if the original scheme is represented by (C1), the modification will be represented by

$$w_t + f(w)_x = \Delta t\{[\beta(w,\lambda)-\gamma(w,\lambda)]w_x\}_x \tag{C3}$$

where $\gamma(w,\lambda)$ is positive. A convenient way to introduce such antidiffusion is by the method of fractional steps:

Step 1: Solve $u_t + f_x = 0$ by using the original scheme. $\tilde{v}_j^{n+1} = (L_1 v^n)_j$.

Step 2. Let D be some finite difference scheme approximating

$$w_t + \Delta t[\gamma(w,\lambda)w_x]_x = 0 . \tag{C4}$$

Take some of the excess of diffusion in $\tilde{v}^{n+1}$ by applying D to it, i.e.

(C5) $$v^{n+1} = D\,\tilde{v}^{n+1} = D\,L_1 v^n$$

Step 2, the antidiffusion, is unstable by itself since it amounts to solving the heat equation backwards. But if

(C6) $$\beta(w,\lambda) - \gamma(w,\lambda) \geq 0$$

then $D \cdot L_1$, the overall scheme is stable. (C6) is a stricter condition on λ than the stability condition for L_1 , i.e. a reduction of the maximal time-step is needed.

We remark that an antidiffusion method cannot all by itself prevent the smearing of a contact discontinuity since $w_t + C\,w_x = \Delta t[\delta(w)w_x]_x$ does not possess a steady progressing solution. Thus the width of the computed discontinuity is still $a \cdot n^{1/2}$; yet the coefficient a can be made small (see (B9)) and thus the width of the computed discontinuity can become acceptable for considerably large time.

The main advantage of the ACM compared to the antidiffusion method is that C is completely stable; therefore no further restrictin on the permissible time-step is needed, and the width of the computed discontinuity, both shock and contact discontinuity, can be made <u>arbitrarily small for all time</u>.

Boris and Book [1] took L_1 to be

(C7) $$\hat{v}_j^{n+1} = (L_1 v^n)_j = v_j^n - \tfrac{1}{2}\mu(v_{j+1}^n - v_{j-1}^n) + (\tfrac{1}{8} + \tfrac{1}{2}\mu^2)(v_{j+1}^n - 2v_j^n + v_{j-1}^n) ,$$

where $\mu = \lambda S$. For $|\mu| \leq \sqrt{3}/2$, L_1 has positive coefficients

and therefore is monotone and L_2-stable. The explicit anti-diffusion

$$(C8)\quad v^{n+1} = (D\,\tilde{v}^{n+1})_j = \tilde{v}_j^{n+1} - \frac{1}{8}(\tilde{v}_{j+1}^{n+1} - 2\tilde{v}_j^{n+1} + \tilde{v}_{j-1}^{n+1})$$

has an amplification factor $g = 1 + \frac{1}{2}\sin^2\frac{\xi}{2}$, $\xi = k\,\Delta x$ for the k-th Fourier component. Thus D is unstable by itself; yet the overall scheme $L_1 \cdot D$ is L_2-stable provided that $|\mu| \leq \frac{1}{2}$ (note the stricter condition). However D is not monotonicity preserving. Boris and Book suggested using the following modification of (C8).

$$(C9)\qquad v_j^{n+1} = \tilde{v}_j^{n+1} - (f_{j+1/2}^c - f_{j-1/2}^c)$$

$$f_c^{j+1/2} = \text{sign}\,(\Delta_{j+1/2})\,\max\{0,\ \min[\Delta_{j-1/2},\ \text{sign}(\Delta_{j+1/2})\},$$

$$\frac{1}{8}|\Delta_{j+1/2}|,\ \Delta_{j+3/2}\cdot\text{sign}\,(\Delta_{j+1/2})]\}$$

where $$\Delta_{j+1/2} = \tilde{v}_{j+1}^{n+1} - \tilde{v}_j^{n+1}\ .$$

This method produces a sharp resolution of shocks and discontinuities of linear equations. The latter fact plus the similarity of (C9) to the Δ-ACF of Section 7 suggests that this modification introduced some element of artificial compression into this scheme. This method although unable to prevent smearing of contact discontinuities for systems, produces much better resolution than that of any standard scheme.

Appendix D. A FORTRAN Subroutine.

In this appendix we give a FORTRAN subroutine performing the operation C_Δ (7.10) in a monotone region, the endpoints of which are J0 and J1. We use a different representation of (7.10) which is computationally more convenient: Let $\tilde{g}_{j+1/2}$ and $\hat{g}_{j+1/2}$ be as in (7.3) and (7.5) We define $g_{j+1/2}$ at the half points by flattening the maxima of $|\hat{g}_{j+1/2}|$, namely

$$(D1)\quad g_{j+1/2} = \begin{cases} \text{sgn}\cdot\max\{|\hat{g}_{j-1/2}|,|\hat{g}_{j+3/2}|\} \\ \qquad \text{if } |\hat{g}_{j+1/2}| < |\hat{g}_{j+1/2}| > |\hat{g}_{j+3/2}| \\ \hat{g}_{j+1/2}\ , \quad \text{otherwise} \end{cases}$$

C_Δ (7.10) can be expressed by $g_{j+1/2}$ in the following way

$$(D2)\quad w_j = (C_\Delta v)_j = v_j - \hat{\lambda}\delta_{j+1}(g_{j+3/2}-g_{j+1/2}) - \hat{\lambda}(1-\delta_{j-1})(g_{j-1/2}-g_{j-3/2})$$

where

$$(D3)\qquad \delta_j = \begin{cases} 1\ , & \text{if } |g_{j+1/2}|-|g_{j-1/2}| < 0 \\ 0\ , & \text{if } |g_{j+1/2}|-|g_{j-1/2}| \geq 0 \end{cases}$$

Observe that $g_{j+1/2}-g_{j-1/2}$ contributes either to w_{j+1} or to w_{j-1}.

The subroutine CDEL operates on the values of the vector V(J) for $J_0 \leq J \leq J_1$; as CDEL terminates the values of $(C_\Delta V)(J)$ are stored in V(J). We denote SGN = sign $[V(J_1)-V(J_0)]$ and ALMD = $\hat{\lambda} \leq 1$ (D2); usually ALMD = 1. In order to save computing time we work directly with differences of the ACF.

```
      SUBROUTINE CDEL(J0,J1,SGN,ALMD,V)
      DIMENSION V(1), ET(  )
      JOP1 = J0+1 $ JIM1 = J1-1
C     Compute η_j = g̃_{j+1/2} - g̃_{j-1/2}   (7.3)
      ET(J0) = V(JOP1) - V(J0) $ ET(J1) = - V(J1)+V(JIM1)
      DO 5 J = JOP1, JIM1
    5 ET(J) = V(J+1) - 2. * V(J) + V(J-1)
C     Compute ĝ_{j+1/2} - ĝ_{j-1/2}   (7.5) and restore in η_j.
      DO 10 J = J0, JIM1
      JP1 = J+1
      IF(ET(JP1)*ET(J).LT.0..AND.ET(JP1)*SGN.GT.0.) 6,10
    6 ETCR = ET(JP1) + ET(J)
      IF (SGN * ETCR) 7,7,8
    7 ET(J) = ETCR $ ET(JP1) = 0. $ GO TO 10
    8 ET(JP1) = ETCR $ ET(J) = 0.
   10 CONTINUE
C     Compute g_{j+1/2} - g_{j-1/2}   (D1)   and restore in η_j.
      DO 15 J = J0, JIM1
      JP1 = J+1
      IF (ET(JP1) * ET(J)) 11,15
   11 ETCR = ET(JP1) + ET(J)
      IF (SGN * ETCR) 13,12,12
   12 ET(J) = ETCR $ ET(JP1) = 0. $ GO TO 15
   13 ET(JP1) = ETCR $ ET(J) = 0.
   15 CONTINUE
```

```
C     Compute  C_Δ V  (D2) using the fact that η_j = g_{j+1/2} - g_{j-1/2}
      contributes either to w_{j+i} = (C_Δ V)_{j+1}  or to w_{j-1} = (C_Δ V)_{j-1}.
      DO 20 J=j0,J1
      IF (ET(J) * SGN) 16,20,17
   16 JM1 = J-1 $ V(JM1) = V(JM1)-ALMD*ET(J) $ GO TO 20
   17 JP1 = j+1 $ V(JP1) = V(JP1)-ALMD*ET(J)
   20 CONTINUE
      RETURN
      END
```

TABLE I

Spread of a Contact Discontinuity Computed by (5.6)

α	Computed Spread	$W(\alpha)$ (5.10)	$\lambda(\alpha)$
0	23 Δx*		0.9
$\frac{3}{8}\sqrt{3}$	9 Δx	9 Δx	0.8
$2.4\sqrt{3}$	$\sim$4,5 Δx	4.1 Δx	0.5

* After 400 Δt .

TABLE II

A Steady Progressing Profile for $\lambda = \frac{1}{2}$, $\hat{\lambda} = 1$ *

$\lambda=\frac{1}{2}$, $\hat{\lambda}=1$ $n=P_{1/2}$	$\lambda=\frac{1}{2}$, $\hat{\lambda}=1$ $n=P_{1/2}+1$	$\lambda=\frac{1}{2}$, $\hat{\lambda}=1$ $n=P_{1/2}+2$	$\lambda=\frac{1}{2}$, $\hat{\lambda}=1$ $n=P_{1/2}+3$	$\lambda=\frac{1}{2}$, $\hat{\lambda}=1$ $n=2P_{1/2}$	$\lambda=\frac{1}{2}$, $\hat{\lambda}=0$ n= 100
1.000	1.000	1.000	1.000	1.000	0.896
⋮	⋮	⋮	⋮	⋮	0.850
					0.850
1.000					0.792
0.999	1.000	1.000			0.792
0.983	0.992	0.997	1.000	1.000	0.722
0.972	0.988	0.995	0.998	0.999	0.722
0.759	0.861	0.922	0.959	0.983	0.642
0.288 →	0.660	0.852	0.938	0.972	0.642
0.	0. →	0.235	0.607 →	0.759	0.553
		0.	0.	0.288 →	0.553
				0.	0.462
					0.462
⋮	⋮	⋮	⋮	⋮	0.371
					0.371
					0.286
					0.286
					0.211
0.	0.	0.	0.	0.	0.211

*Arrow indicates the exact location of the discontinuity.

TABLE III

Profiles of the ACM for a Contact Discontinuity *

$\lambda=0,\ \hat{\lambda}=1,$ $P_0=1$	$\lambda=\frac{1}{4},\ \hat{\lambda}=1$ $P_{1/4}=8$	$\lambda=1/2,\ \hat{\lambda}=1$ $P_{1/2}=4$	$\lambda=3/4,\ \hat{\lambda}=1$ $P_{3/4}=8$	$\lambda=1,\ \hat{\lambda}=1$ $P_1=1$
1.000	1.000	1.000	1.000	1.000
	1.000	0.999	1.000	
	0.989	0.983	0.992	
	0.982	0.972	0.975	
→ 1.000	→ 0.737	→ 0.759	→ 0.816	→ 1.000
0.	0.293	0.288	0.217	0.
0.	0.	0.	0.	0.

*Arrows indicate the exact location of the discontinuity.

TABLE IV

Profiles of the ACM for the Shock (9.5) *

$\lambda=0, \hat{\lambda}=1$ $P_0=1$	$\lambda=\frac{1}{4}, \hat{\lambda}=1$ $P_{1/4}=8$	$\lambda=\frac{1}{2}, \hat{\lambda}=1$ $P_{1/2}=4$	$\lambda=\frac{3}{4}, \hat{\lambda}=1$ $P_{3/4}=8$	$\lambda=1, \hat{\lambda}=1$ $P_1=1$	$\lambda=\frac{1}{2}, \hat{\lambda}=\frac{1}{2}$ $P_{1/2}=4$	$\lambda=\frac{1}{2}, \hat{\lambda}=0$ $P_{1/2}=4$
1.000	1.000	1.000	1.000	1.000	0.996	0.896
.					0.989	0.896
.	1.000	1.000	1.000		0.969	0.740
	0.999	0.997	0.999		0.916	0.740
1.000	0.934	0.959	0.973	1.000	0.785	0.494
→ 0. →	→ 0.567 →	→ 0.543 →	→ 0.528 →	→ 0. →	→ 0.525 →	0.494
	0.	0.	0.		0.214	0.254
.					0.073	0.254
.					0.024	0.107
					0.008	0.107
0.	0.	0.	0.	0.	0.002	0.040

*Arrows indicate the exact location of the shock.

TABLE V

A Profile of the ACM for the Contact Discontinuity of the Polytropic Gas (13.15a)

Density ρ	Momentum m	Energy E	Velocity u	Pressure p
2.0000	4.0000	10.0000	2.0000	2.4000
1.9999	3.9997	9.9997		
1.9998	3.9996	9.9996		
1.9966	3.9931	9.9931		
1.9945	3.9890	9.9890		
1.9224	3.8449	9.8449		
1.8518	3.7035	9.7035		
1.2351	2.4701	8.4701		
1.0000	2.0000	8.0000	2.0000	2.4000

TABLE VI

Comparison between Profiles of the LF Scheme

and the Corresponding Viscous Profiles

LF Profile $\lambda = 0.2$	Viscous Profile $\lambda = 0.2$	LF Profile $\lambda = 0.4$	Viscous Profile $\lambda = 0.4$	LF Profile $\lambda = 0.6$	Viscous Profile $\lambda = 0.6$	LF Profile $\lambda = 0.8$	Viscous Profile $\lambda = 0.8$
0.947	0.950	0.999	0.999				
0.922	0.926	0.997	0.998				
0.886	0.890	0.994	0.995	1.000			
0.833	0.838	0.985	0.990	0.999	1.000		
0.761	0.766	0.966	0.975	0.997	0.997	1.000	
0.663	0.667	0.923	0.938	0.988	0.993	0.999	1.000
0.535	0.538	0.833	0.853	0.952	0.978	0.991	0.996
0.378	0.380	0.658	0.677	0.831	0.881	0.929	0.990
0.196	0.197	0.372	0.383	0.519	0.555	0.637	0.729
0.	0.	0.	0.	0.	0.	0.	0.
-0.196	-0.197	-0.372	-0.383	-0.519	-0.555	-0.637	-0.729
-0.378	-0.380	-0.658	-0.677	-0.831	-0.881	-0.929	-0.990
-0.535	-0.538	-0.833	-0.853	-0.952	-0.978	-0.991	-0.996
-0.663	-0.667	-0.923	-0.938	-0.988	-0.993	-0.999	-1.000
-0.761	-0.766	-0.966	-0.975	-0.997	-0.997	-1.000	
-0.833	-0.838	-0.985	-0.990	-0.999	-1.000		
-0.886	-0.890	-0.994	-0.995	-1.000			
-0.922	-0.926	-0.997	-0.998				
-0.947	-0.950	-0.999	-0.999				

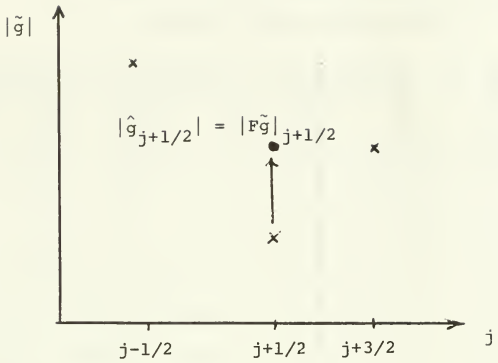


Figure 1. Illustration of (7.5)

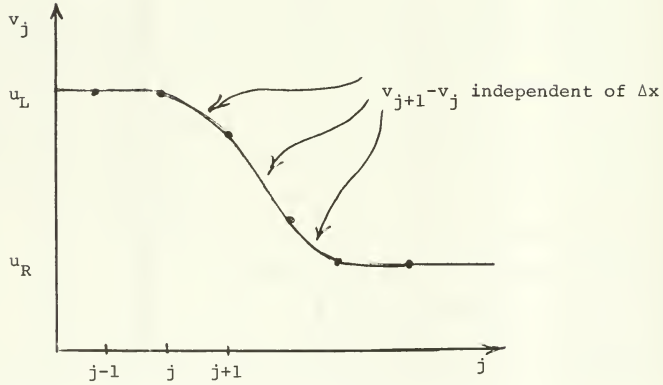


Figure 2. Illustration of the Need to Define $\tilde{g}_{j+1/2}$ at the Midpoints.

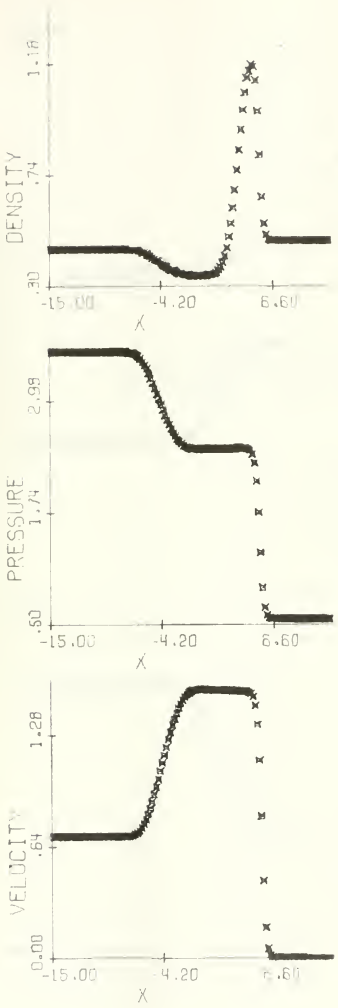
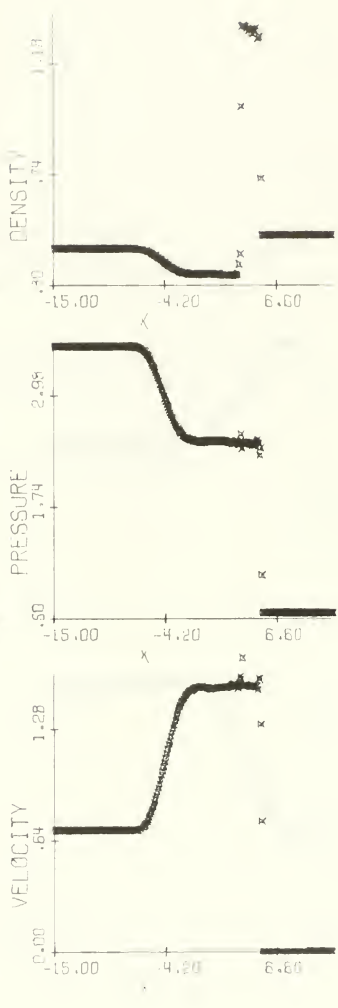


Fig. 3a
The Riemann problem by the LF scheme - 100 Δt.



Fib. 3b
The Riemann problem by the ACM - 100 Δt.

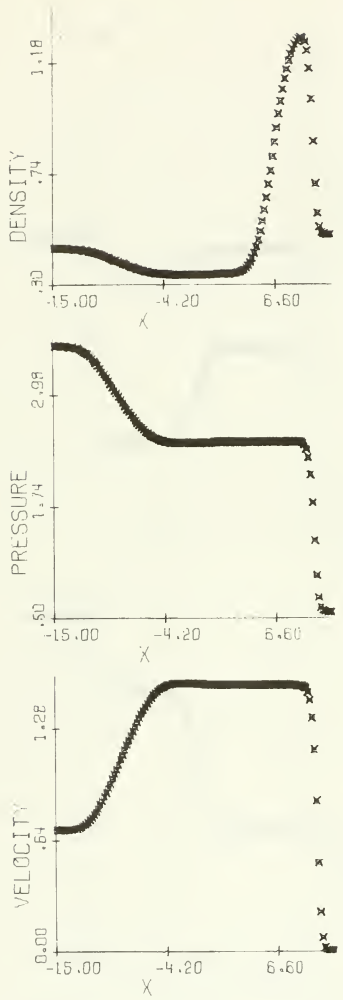


Fig. 3c
The Riemann problem by the RF scheme - 200 Δt.

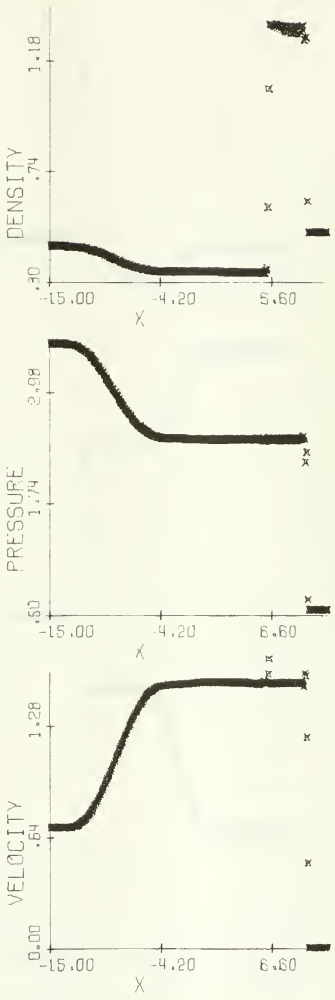


Fig. 3d
The Riemann problem by the ACM - 200 Δt.

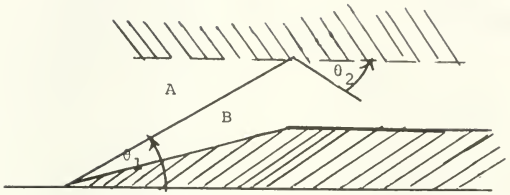


Fig. 4a. The Oblique Reflection.

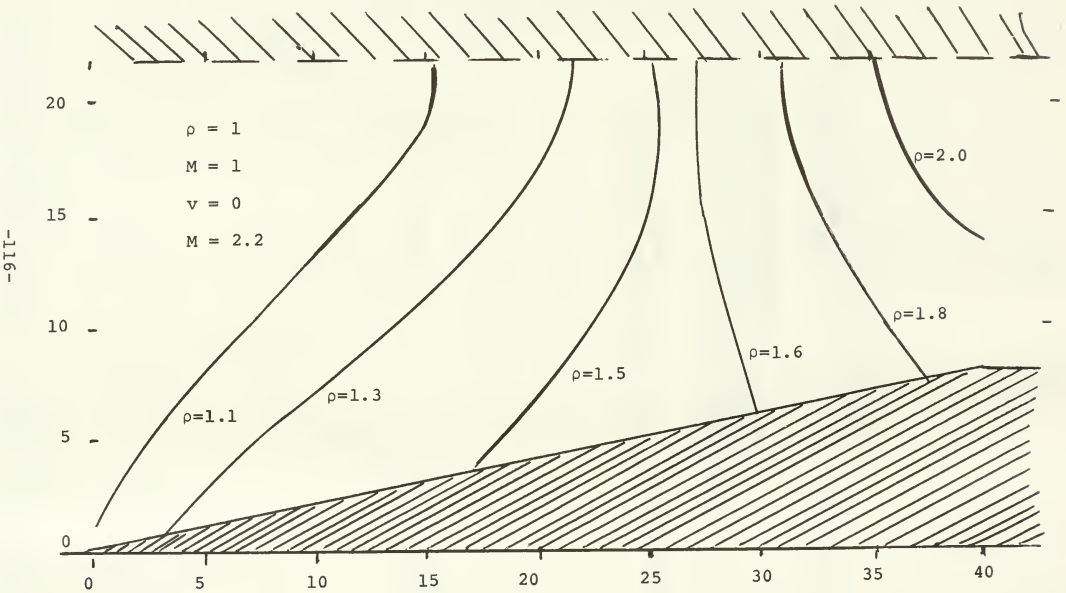


Fig. 4b. Oblique Reflection by the LF Scheme ($\hat{\lambda}$ = 0) -- Density of Steady State.

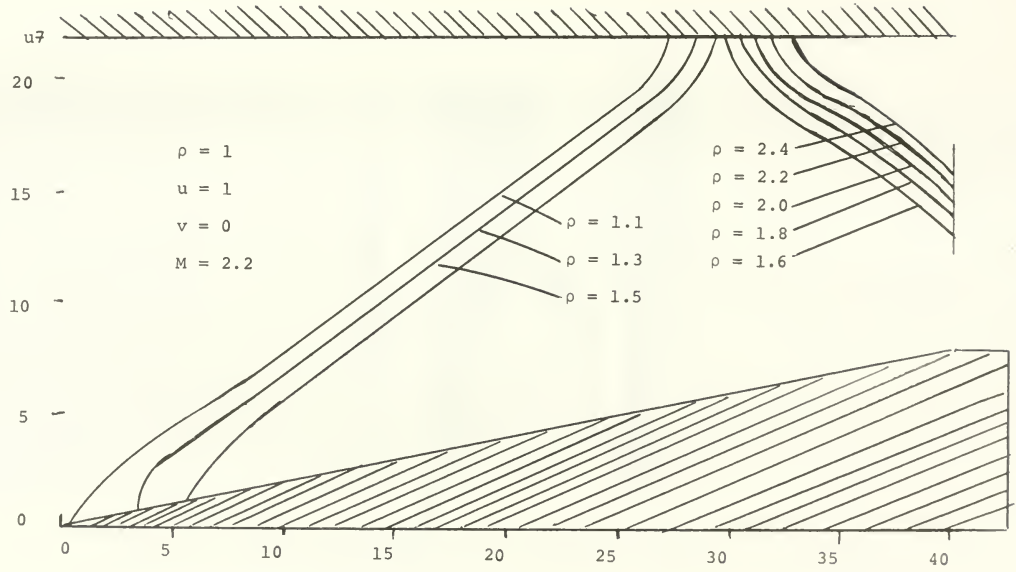


Fig. 4c. Oblique Reflection by the ACM ($\hat{\lambda}$ = 0.9) -- Density of Steady State.

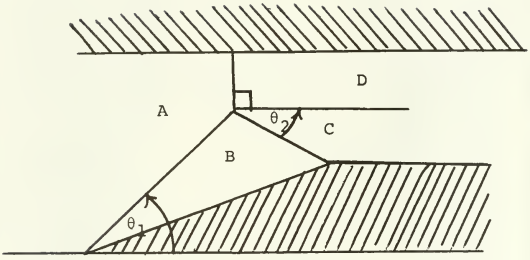


Fig. 5. The Mach Reflection.

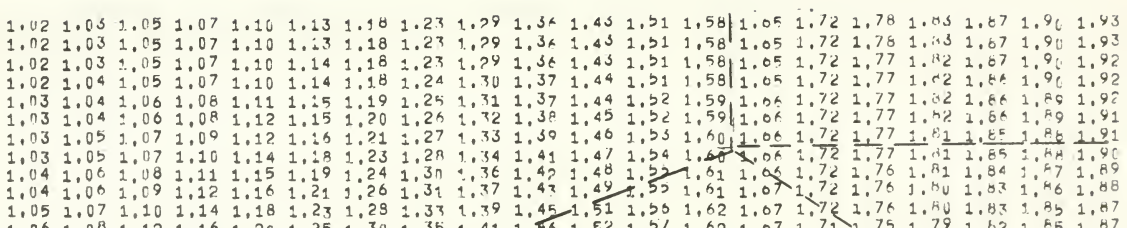


Figure 6a. Mach Reflection by the LF Scheme - Density ρ at T = 1000 Δt.

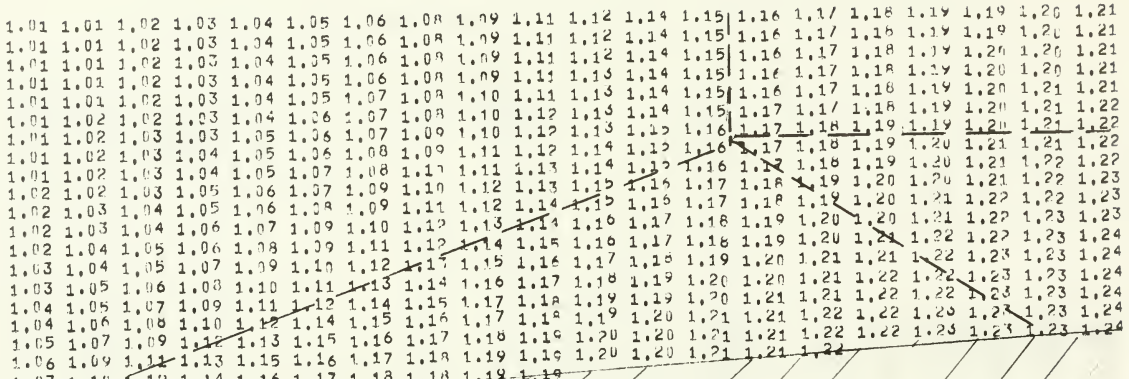


Figure 6b. Mach Reflection by the LF-Scheme - Momentum in x-direction m at T = 1000 Δt.

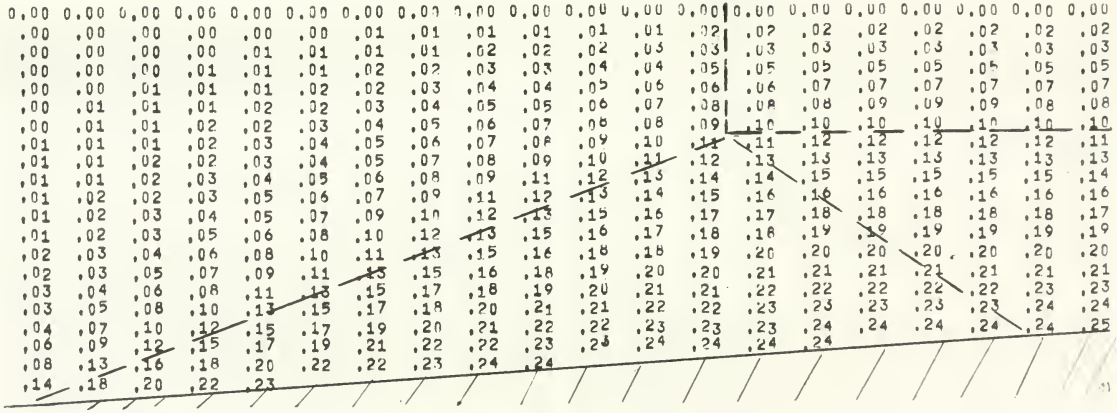


Figure 6c. Mach Reflection by the LF Scheme - Momentum in y-direction n at T = 1000 Δt.

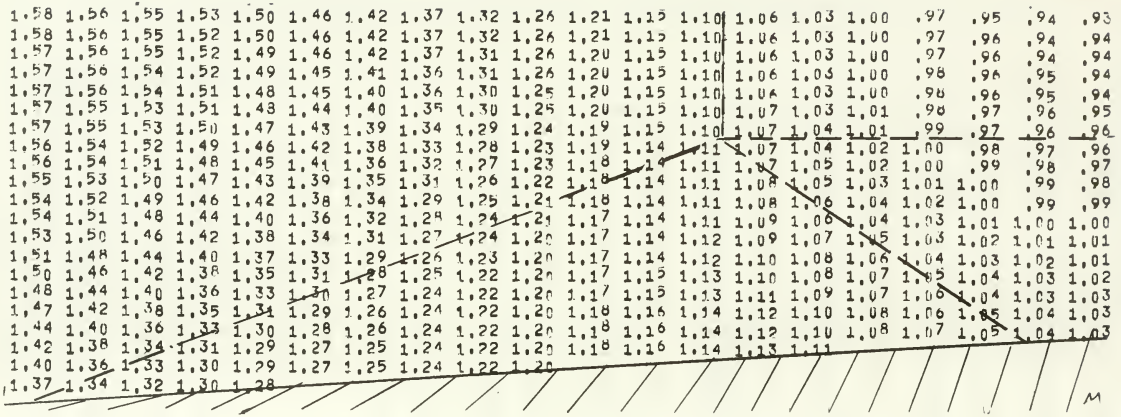


Figure 6d. Mach Reflection by the LF Scheme - Mach Number M at T = 1000 Δt.

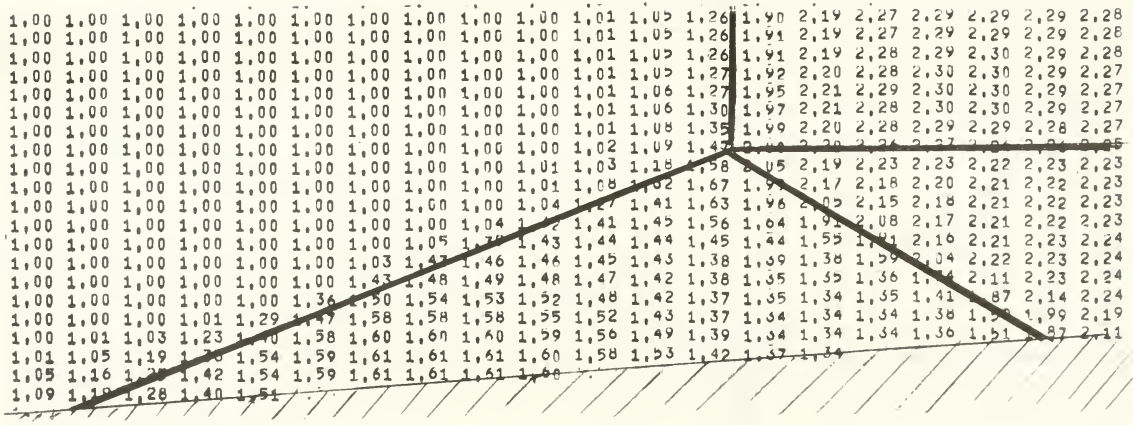


Figure 7a. Mach Reflection by the ACM -- Density ρ at T = 1000 Δt

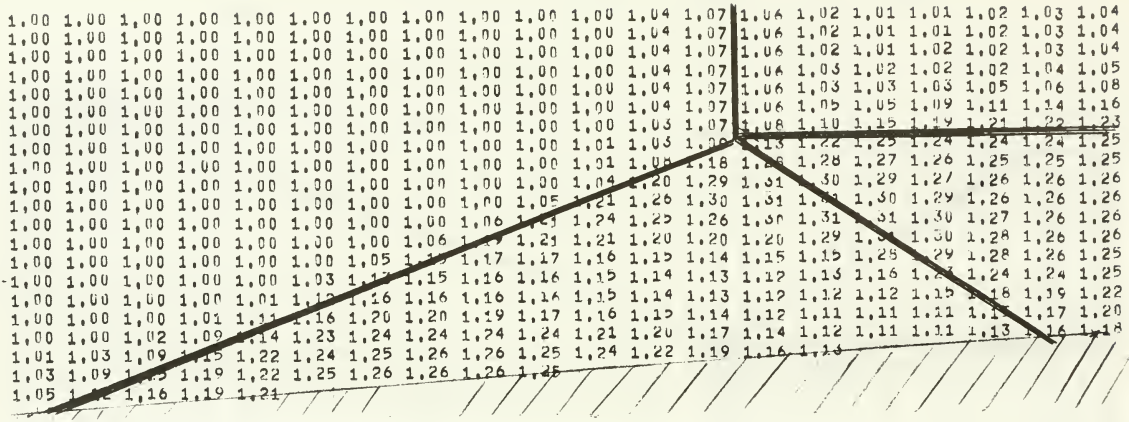


Figure 7b. Mach Reflection by the ACM - Momentum in x-direction m at T = 1000 Δt

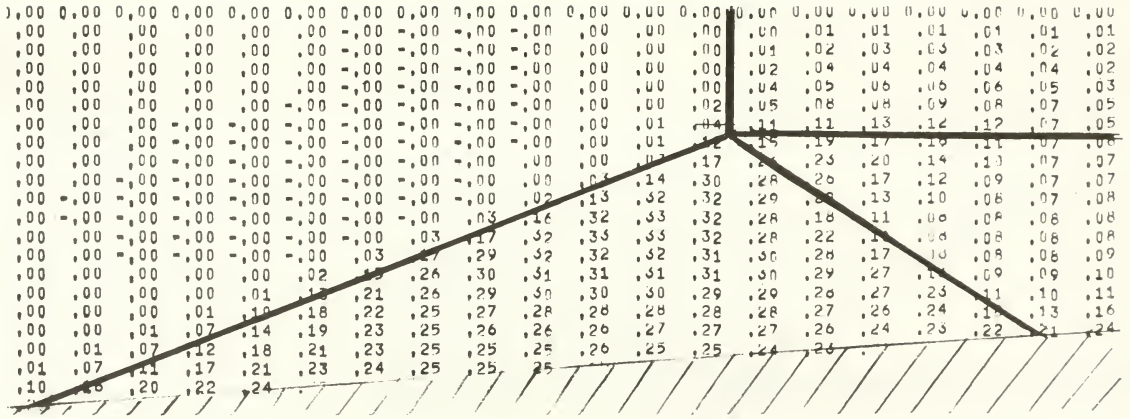


Figure 7c. Mach Reflection by the ACM - Momentum in the y-direction n at T = 1000 Δt

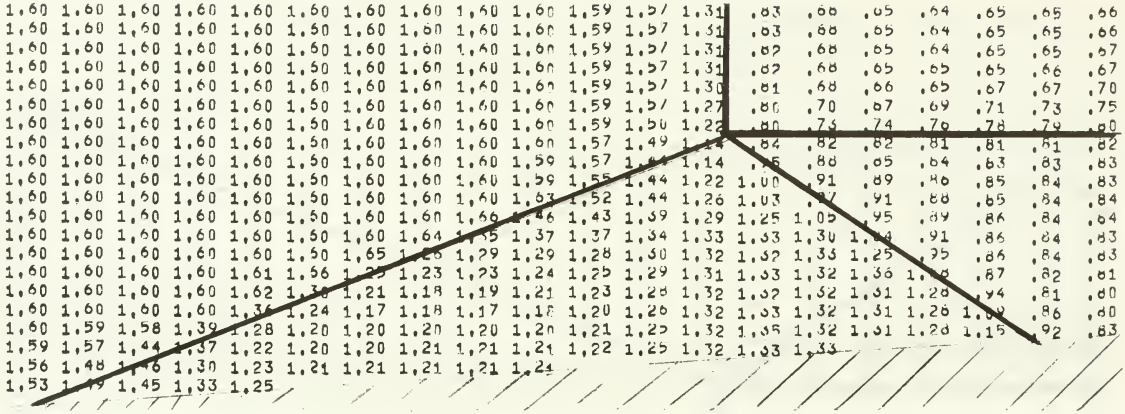


Figure 7d. Mach Reflection by the ACM - Mach Number M at T = 1000 Δt.

This report was prepared as an account of Government sponsored work. Neither the United States, nor the Commission, nor any person acting on behalf of the Commission:

A. Makes any warranty or representation, express or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

B. Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used in the above, "person acting on behalf of the Commission" includes any employee or contractor of the Commission, or employee of such contractor, to the extent that such employee or contractor of the Commission, or employee of such contractor prepares, disseminates, or provides access to, any information pursuant to his employment or contract with the Commission, or his employment with such contractor.

MAY 20 1974
DATE DUE
SEP 16 1981
OCT
MAR 11 1982
JAN 21 1987
FEB 24 1987
MAR 09 1987
MAR 09 1987
GAYLORD
PRINTED IN U.S.A.

COO-3077-50 c.2

Harten

The method of artificial
com[illegible] Part 1.

COO-3077-50 c.2

AUTHOR Harten

TITLE The method of artificial
compression. Part 1.

DATE DUE | BORROWER'S NAME

SAMPERT

N.Y.U. Courant Institute of
Mathematical Sciences
251 Mercer St.
New York, N. Y. 10012